



UNIVERSIDAD NACIONAL DE COLOMBIA

Metodología estadística para el análisis de datos funcionales cerebrales: Una aproximación con potenciales evocados.

Jeimy Paola Aristizabal Rodríguez

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá, Colombia
2011

Metodología estadística para el análisis de datos funcionales cerebrales: Una aproximación con potenciales evocados

Jeimy Paola Aristizabal Rodríguez

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Estadística

Director(a):

Ph.D., Doctor en Estadística, Ramón Giraldo Henao

Línea de Investigación:

Análisis de datos funcionales

Universidad Nacional de Colombia

Facultad de Ciencias, Departamento de Estadística

Bogotá, Colombia

2011

A ...

Dios, por haberme dado la oportunidad de
recorrer este camino.

A ...

Mi padre por ser siempre mi inspiración.

A ...

Mi madre por su esfuerzo, dedicación y
tenacidad.

A ...

Mis abuelas por su eterno amor.

A ...

Mis hermanos Andrea y Jorge por los miles
de momentos compartidos.

A ...

Edisson por su infinita paciencia y cariño.

A ...

La memoria de mi tío, mis abuelos y del
profesor Leonardo Bautista.

Agradecimientos

Es imposible recordar a todas las personas que en algún momento pusieron su granito de arena para éste logro. Empiezo agradeciendo a Dios por sus infinitas bendiciones, a mis padres porque sé que gracias a su esfuerzo soy lo que soy. A mis hermanitos por su amor y entusiasmo, a la familia Sanabria Villaizan por su incondicional apoyo. A Oscar, Aleja y Dago por sus eternas tardes de estudio. A Tatiana y a Claudia por su compañerismo. Al profesor Oscar Soto por sus cientos de enseñanzas, al profesor Pedro Nel Pacheco por su paciencia y colaboración y a la profesora Martha Bohórquez por su alegría característica. Finalmente a mi director de tesis Ramón Giraldo por darme la oportunidad de aprender y a Jorge Mateu por toda su disposición.

“El éxito no es para siempre y el fracaso no es el final, lo que cuenta es el valor de seguir adelante...”

Resumen

El estudio del comportamiento, de los procesos de cognición y de las causas cerebrales por las cuales se padece una determinada enfermedad, ha sido un tema de estudio en neurología y en otras áreas del conocimiento. En éste sentido la estadística ha aportado procedimientos que permiten detectar las fuentes de tales procesos y causas. En muchos casos se asume independencia, desconociéndose la estructura espacial inherente a este tipo de datos. En éste trabajo se tiene en cuenta la correlación espacial dentro de los análisis. En particular se avalúa la dependencia espacial entre curvas de potenciales evocados y se involucra ésta misma en pruebas ANOVA funcionales que tienen como propósito detectar diferencias entre géneros o tipos de estímulo.

Se estudian datos de la respuesta bio-eléctrica del cerebro cuando es estimulado por palabras con carga emocional positiva, negativa o neutra. Se tienen curvas de la medición de potenciales evocados obtenidas en 19 electrodos en cada uno de 48 individuos (30 hombres y 18 mujeres). En una primera etapa se pre-procesan los datos suavizándolos a través de bases de B-splines. Posteriormente se realizan pruebas ANOVA funcionales para detectar diferencias entre géneros (fijando un electrodo) o entre tipos de estímulos (fijando un individuo). En este último caso, se acude al uso de geoestadística funcional para estimar la estructura de correlación espacial y ésta es involucrada en la estimación de los parámetros del ANOVA funcional y en la aplicación de las respectivas pruebas. Al fijar un electrodo, los resultados sugieren que no existen diferencias significativas ni entre géneros ni entre tipos de estímulos. Fijando un paciente, en algunos, los estímulos presentan diferencias significativas y en otros no.

Palabras clave: Correlación espacial, covariograma, datos cerebrales, datos funcionales, electrodo, electroencefalograma, semivariograma, traza-variograma..

Abstract

The study of the behavior, the cognitive processes, and the brain causes why someone has a disease, have been subject of study in neurology and other areas. In this sense the statistics analysis has given procedures for detecting sources of such processes and causes. The methodologies for analyzing brain data, usually do not take into account the spatial correlation structure present in this type of data. In this paper we propose and implement a methodology for modeling brain data by using functional geostatistical methods in order to estimate the spatial correlation structure between electroencephalogram curves. In addition, we propose the use of functional analysis of variance which involve the spatial correlation structure. This analysis can be used to detect differences among different types of impulses. The results obtained with a real data set show that in men there are differences in brain electrical activity between the impulses, while in women no statistically significant differences were observed.

Keywords: Brain data, EEG, electrode, functional data, Spatial correlation, trace-variogram.

Contenido

Agradecimientos	vii
Resumen	ix
Introducción	5
1. Datos cerebrales	1
1.1. Cartografía del cerebro	1
1.2. Electroencefalograma	2
1.3. Potenciales evocados (P300)	3
2. Marco teórico	6
2.1. Geostatística	6
2.1.1. Geoestadística univariada	6
2.1.2. Geoestadística multivariada	8
2.1.3. Modelo lineal de correogionalización (MLC)	9
2.2. Análisis de datos funcionales (ADF)	10
2.2.1. Definiciones	10
2.2.2. Suavizado de funciones mediante funciones B-splines	11
2.2.3. Herramientas descriptivas del Análisis de datos Funcionales	12
2.3. Modelos lineales funcionales (Regresión Funcional)	12
2.3.1. Propuesta de Ramsay and Silverman (2005)	13
2.3.2. Propuesta de Cuevas, Febrero and Fraiman (2004)	15
2.3.3. Prueba de Shen and Faraway (2004)	16
2.3.4. Prueba de Cuesta-Albertos and Febrero-Bande (2007)	18
2.4. Geoestadística de datos funcionales	21
3. ANOVA funcional espacialmente correlacionado	23
3.1. Transformación de los datos a datos funcionales	23
3.2. Estimación de la estructura de dependencia espacial	23
3.3. Estimación de los parámetros del ANOVA funcional. Extensión de la propues- ta de Ramsay and Silverman (2005)	25
3.4. Extensión de la prueba de Shen and Faraway (2004)	26
3.5. Extensión de la Prueba de Cuesta-Albertos and Febrero-Bande (2007)	28

3.6. Resumen	29
4. Aplicación	31
4.1. Descripción de la base de datos	31
4.1.1. Sistema de referencia 10-20	31
4.2. Análisis de varianza funcional sin estructura de dependencia espacial.	32
4.2.1. Análisis de varianza funcional clásico para los hombres	33
4.2.2. Análisis de varianza para las mujeres	36
4.3. Análisis de varianza a dos vías de clasificación	40
4.4. Análisis de varianza con datos correlacionados espacialmente	42
4.4.1. Paciente hombre	43
4.4.2. Paciente mujer	46
5. Conclusiones	52
6. Trabajo futuro	53
A. Funciones en R para realizar ANOVAF con errores independientes	54
A.1. Código para ANOVAF de los pacientes de un género determinado en un elec- trodo determinado	54
A.2. Código para ANOVAF en un electrodo determinado (a dos vías)	66
B. Funciones en R para realizar ANOVAF con errores correlacionados	79
B.1. Código para ANOVAF de un paciente determinado considerando las curvas en los 19 electrodos	79
Bibliografía	91

Lista de Tablas

3.1. Resumen de aplicabilidad de las pruebas clásicas y de las pruebas propuestas	30
4.1. Prueba F funcional de Shen and Faraway (2004) para el ANOVAF propuesto en la ecuación (4.5).	42
4.2. Parámetros estimados del modelo de las traza variogramas mediante MLC para el paciente hombre identificado con el número 5.	45
4.3. Pruebas ANOVAF para el paciente hombre identificado con el número 5.	46
4.4. Parámetros estimados del modelo de las traza variogramas mediante MLC para la paciente mujer identificada con el número 2.	47
4.5. Pruebas ANOVAF para la paciente mujer identificada con el número 2.	48

Lista de Figuras

4.1. Coordenadas en el sistema 10-20.	32
4.2. Suma de cuadrados del error (SSE) obtenida por validación cruzada en función del número de funciones de la base.	34
4.3. Curvas suavizadas de los 30 hombres para el estímulo positivo (rojo) y para el estímulo negativo (azul) en la ubicación del electrodo C3 (Izquierda). Curva media y de desviación estándar (Derecha).	35
4.4. Parámetros funcionales estimados para el análisis de varianza funcional de los hombres en la ubicación del electrodo C3.	36
4.5. Estimación de la curva media respuesta para los hombres con el estímulo positivo (rojo) y estímulo negativo (azul).	37
4.6. Prueba F para la diferencia entre los estímulos en cada tiempo (Izquierda) y prueba funcional de Cuevas and Fraiman (2004) (Derecha).	37
4.7. Curvas suavizadas de las 18 mujeres para el estímulo positivo (rojo) y para el estímulo negativo (azul) en la ubicación del electrodo C3.	38
4.8. Parámetros funcionales estimados para el análisis de varianza funcional de las mujeres en el electrodo C3.	39
4.9. Estimación de la curva media respuesta para las mujeres con el estímulo positivo (rojo) y estímulo negativo (azul).	40
4.10. Prueba F para la diferencia entre los estímulos (Izquierda) y prueba funcional de Cuevas and Fraiman (2004) (Derecha)	40
4.11. Estimación de la curva respuesta media para todos los pacientes en la ubicación del electrodo C3.	41
4.12. Parámetros funcionales estimados para el ANOVA funcional de los pacientes en el electrodo C3	42
4.13. Estimación de la curva respuesta para los hombres con estímulo positivo (rojo), hombres con estímulo negativo (azul), mujeres con estímulo positivo (verde) y mujeres con estímulo negativo (morado) en el electrodo C3.	43
4.14. Suavizamiento de funciones a través de B-splines para el paciente hombre identificado con el número 5 (Izquierda). Curva media de la respuesta bioeléctrica del cerebro al ser estimulado con palabras positivas (rojo) y negativas (verde) (Derecha).	44

4.15. Estimaciones de las traza variogramas y de la traza variograma cruzada para el paciente hombre identificado con el número 5. Los puntos rojos presentan la traza variograma bin y los negros la traza variograma cloud.	45
4.16. Ajuste del modelo exponencial a las traza variogramas del paciente hombre identificado con el número 5.	46
4.17. Estimación de los parámetros funcionales involucrando la estructura de correlación espacial para el paciente hombre identificado en el número 5 (Izquierda). Estimación de las medias funcionales (Derecha).	47
4.18. Estimación de los parámetros funcionales sin involucrar la estructura de correlación espacial para el paciente hombre identificado en el número 5 (Izquierda). Estimación de las medias funcionales (Derecha).	48
4.19. Suavizamiento de funciones a través de B-splines para la paciente mujer identificada con el número 2 (Izquierda). Curva media de la respuesta bio-eléctrica del cerebro al ser estimulado con palabras positivas (rojo) y negativas (verde) (Derecha).	49
4.20. Estimaciones de las traza variogramas y de la traza variograma cruzada para la paciente mujer identificada con el número 2. Los puntos rojos presentan la traza variograma bin y los negros la traza variograma cloud.	49
4.21. Ajuste del modelo exponencial a las traza variogramas de la paciente mujer identificada con el número 2.	50
4.22. Estimación de los parámetros funcionales involucrando la estructura de correlación espacial para la paciente mujer identificada en el número 2 (Izquierda). Estimación de las medias funcionales (Derecha).	50
4.23. Estimación de los parámetros funcionales sin involucrar la estructura de correlación espacial para la paciente mujer identificada en el número 2 (Izquierda). Estimación de las medias funcionales (Derecha).	51

Introducción

El estudio del comportamiento, de los procesos de cognición y de las causas cerebrales por las cuales se padece una determinada enfermedad, ha sido de interés en neurología y otras áreas del conocimiento. Todas ellas aportan metodologías y procedimientos que permiten detectar las fuentes de tales procesos y causas.

El sistema nervioso central está constituido prácticamente en su totalidad por células excitables que presentan actividad eléctrica espontánea o provocada a consecuencia de los cambios de potencial de las distintas unidades neuronales (Duffy and Burchfiel, 1981).

Stuart and Robert (2008), mostraron que había variación del potencial en la superficie del cerebro al recibir un estímulo. Posteriormente se mostró que dependiendo de la zona se podían provocar variaciones en la actividad eléctrica a través de la estimulación de diversos órganos o miembros, conociéndose esto como potenciales evocados.

La actividad eléctrica de las neuronas se registra con tres tipos de electrodos: superficiales, corticales y de profundidad. Cuando se utilizan electrodos superficiales, el registro es llamado electroencefalograma (EEG) y cuando los electrodos se ubican en la corteza cerebral, el registro es denominado electrocorticograma (ECoG). Los electrodos pueden penetrar más profundamente en el cerebro para obtener los llamados “registros de profundidad”. Los potenciales registrados mediante estas tres técnicas representan la actividad de numerosas neuronas en las que se produce una fluctuación en el potencial trans-membrana (Stuart and Robert, 2008).

El registro del EEG se realiza con distintos tipos de electrodos distribuidos según un mapa normalizado por la Federación de Electroencefalografía y Neurofisiología Clínica (FENC), con base en el sistema de coordenadas 10-20, que se fundamenta en determinar las coordenadas de cada punto de registro mediante porcentajes (10-20) de la distancia existente entre puntos del cráneo bien definidos (Chatrian, 1985). Esto permite el estudio de la misma región independientemente de las dimensiones y de la forma del cráneo de un paciente.

En el primer Congreso Internacional de EEG, que se celebró en Londres en 1947, se reconoció que era necesario un método estándar de la colocación de los electrodos utilizados en el electroencefalograma (EEG). Varias discusiones con respecto al tema dieron lugar a

la definición del sistema de electrodos 10-20 (Hasper, 1958). Desde entonces, éste sistema de electrodos se ha convertido en el estándar para los clínicos y para el estudio de potenciales relacionados con eventos (potenciales evocados) en entornos no clínicos. Sin embargo, el avance de los sistemas multi-canal de EEG y de los métodos topográficos para estudiar potenciales evocados relacionados con eventos, hizo necesaria la normalización de un mayor número de canales. Por lo tanto una extensión del sistema original 10-20, que aumenta el número de electrodos de 21 hasta 74 se propuso en años posteriores (Chatrian, 1985).

Los registros de EEG pueden ser “monopolares” o “bipolares”. Los monopulares se efectúan por medio de un electrodo activo en el cuero cabelludo y un electrodo indiferente en un tejido a distancia, generalmente ubicado en los lóbulos de las orejas. Los bipolares se hacen con dos electrodos aplicados en sitios activos del cuero cabelludo. El potencial registrado entre ellos en cualquier instante es la suma algebraica del voltaje que aparece debajo de cada electrodo (Sánchez, 1992).

Desde la física y la estadística se han hecho contribuciones importantes para el análisis de ésta información. Un ejemplo claro de este avance se tiene en el estudio de potenciales evocados para detectar diferencias en las respuestas bio-eléctricas del cerebro que estén relacionadas con la producción de procesos comportamentales, psicológicos o físicos, tales como percepción, motivación, aprendizaje o patologías como por ejemplo la epilepsia. Aunque éstos fenómenos son complejos y pueden ser estudiados desde otras perspectivas como la neuroanatomía, neuroquímica o genética, un análisis estadístico en potenciales evocados es muy valioso como puerta de entrada a una explicación fisiológica del comportamiento y al entendimiento de la fisiología o funcionamiento de distintas partes del cerebro y de estructuras implicadas en la percepción y cognición.

El interés por este tipo de investigación es debido a varias razones: La sensibilidad de técnicas de mapeado cerebral a variaciones en procesos psicológicos, la posibilidad de efectuar registros fisiológicos de forma no invasiva y sin problemática ética y el costo relativamente asequible y su mayor resolución temporal comparado con otras técnicas de mapeado cerebral (como la Resonancia Magnética Funcional, o la Tomografía de Emisión de Positrones). Sin embargo, su interpretación visual provee una localización imprecisa indicando, en el mejor de los casos, cuál es el lóbulo involucrado en un determinado proceso (Ashburner and Friston, 1997).

Los datos de potenciales evocados medidos a través del electroencefalograma (EEG) no tienen la estructura comúnmente manejada en estadística, es decir, una respuesta escalar para cada unidad muestral, sino que para cada una de ellas (electrodos en este caso) se cuenta con M respuestas a través del tiempo (dato funcional). El propósito es brindar a investigadores de diferentes campos herramientas que permitan evaluar y predecir la actividad cerebral y

determinar el efecto de covariables en ésta misma.

La posibilidad de obtener para un mismo individuo un gran número de registros de la misma característica ha propiciado la aparición de nuevos métodos estadísticos que facilitan el trabajo con grandes volúmenes de información. Entre ellos se encuentran las técnicas no paramétricas de suavizado (Delicado, 2007) y el análisis de datos funcionales (ADF).

El ADF (y de curvas en particular), es una nueva visión que considera el caso en el que el número de medidas repetidas es muy grande. Su reciente crecimiento involucra procedimientos de muchas áreas de la estadística. Cabe destacar las contribuciones en análisis exploratorio y descriptivo de datos (Ramsay, 2005), modelos lineales (Cardot and Sarda, 1999; Malfait and Froda, 2000), modelos lineales generalizados (Escabias and Valderrama, 2004), análisis de varianza (Cuevas, Febrero and Fraiman, 2004; Delicado, 2007; Shen and Faraway, 2004), métodos no paramétricos (Ferraty and Vieu, 2006), datos longitudinales (Yao and Wang, 2005) y técnicas multivariadas como componentes principales (Pezulli and Silverman, 1993; Silverman, 1995), correlación canónica (Leurgans and Silverman, 1993), análisis discriminante (Ferraty and Vieu, 2003) y análisis cluster (Clarkson and Ramsay, 2005). En la mayoría de los casos se supone que las curvas son independientes, situación que en algunas áreas del conocimiento no se cumple. En muchas situaciones experimentales el supuesto de independencia puede ser inválido, por ejemplo cuando las funciones son evaluadas en diferentes puntos del tiempo (temporalmente correlacionadas) o del espacio (espacialmente correlacionadas). En éste caso, algunos de los procedimientos pierden propiedades importantes, razón por la cual es necesario extender dichas metodologías considerando este tipo de correlación. Se destaca que en la literatura para datos funcionales correlacionados temporalmente, Yamanishi y Tanaka (2003) desarrollaron un modelo de regresión que tiene en cuenta la relación entre las variables en el tiempo y en el espacio, combinando metodologías de regresión ponderada geográficamente (Brunsdon, Fotheringham and Charlton, 1998) y de regresión funcional múltiple (Ramsay and Silverman, 2005) y Baladandayuthapani, Mallick, Hong, Lupton, Turner, y Carroll (2008) muestran una alternativa para analizar un diseño experimental con una variable funcional respuesta correlacionada, a través de procedimientos jerárquicos Bayesianos.

La información generada por el cerebro puede ser tratada desde el contexto de datos funcionales. En este trabajo se propone una metodología con este fin. En particular se estudian datos correspondientes a la respuesta bio-eléctrica del cerebro al ser estimulado con palabras de diferente carga emocional (positiva, negativa o neutra). En cada individuo se tienen 19 ubicaciones (electrodos) georeferenciadas bajo un sistema normalizado por la FENC (Chatrian, 1985), con base en el sistema de coordenadas 10-20. Esta información es representada

mediante curvas a través del uso de un sistema de bases B-splines. Los parámetros óptimos del suavizado se obtienen por validación cruzada no paramétrica. Una vez obtenidas dichas curvas (objetos funcionales), se realiza un análisis de varianza funcional con la información de todos los pacientes en un electrodo previamente fijado, con el fin de determinar si existen diferencias significativas en la actividad cerebral dado el tipo de palabra por la cual es estimulado el paciente. Finalmente, se considera un modelo en el que se incluye la respuesta cerebral de un paciente en las 19 ubicaciones (electrodos), extendiendo la metodología de análisis de varianza funcional al caso en el que se presenta una estructura de correlación espacial. La dependencia espacial es estimada a través de las funciones traza-variograma (Giraldó, 2009) y traza variograma cruzada (propuesta en éste trabajo). Se emplea el método de mínimos cuadrados generalizados para estimar los parámetros del modelo ANOVA funcional.

La tesis está organizada de la siguiente forma:

En el capítulo 1 se precisan los conceptos cerebrales básicos necesarios para entender la aplicación a potenciales evocados y se presentan algunas herramientas estadísticas que hasta el momento se han utilizado en el contexto cerebral. En el capítulo 2, se establece el marco teórico usado como base para realizar la propuesta metodológica presentada en el capítulo 3. En dicho marco se involucran conceptos de análisis de datos funcionales, geoestadística funcional y modelos de regresión y análisis de varianza funcional. Como se mencionó anteriormente en el capítulo 3 se realiza la propuesta metodológica para realizar un análisis de varianza funcional con correlación espacial. Se explican las etapas para estimar la matriz de varianza y covarianza espacial. Esta matriz es incluida en el proceso de estimación de los parámetros funcionales y en el cálculo de las estadísticas de prueba de los test aplicados. Finalmente en el capítulo 4 se realiza una aplicación de la metodología propuesta al caso de mediciones de electroencefalogramas. En la primera parte no se considera correlación espacial, dado que se fija un electrodo en particular (se aplican los procedimientos de análisis de varianza funcional clásicos, mostrados en el capítulo 2). En la segunda aplicación, se fija un paciente y con la información obtenida se realiza un análisis de varianza funcional en presencia de correlación espacial llevando a cabo cada una de las etapas presentadas en el capítulo 3.

En el apéndice se muestra el código utilizado. Este puede ser descargado de la página [http : //www.docentes.unal.edu.co/rgiraldoh/docs/](http://www.docentes.unal.edu.co/rgiraldoh/docs/). Sin embargo, por restricciones con el uso de la información analizada los datos no están disponibles en dicha dirección.

1. Datos cerebrales

En la primera parte de este capítulo se presenta un resumen de las herramientas estadísticas que se han utilizado para analizar datos cerebrales y como éstas han progresado a través del tiempo con el desarrollo y adaptación de diversos métodos. Posteriormente se precisan los conceptos de electroencefalograma y potenciales evocados necesarios para entender la aplicación de la metodología propuesta.

1.1. Cartografía del cerebro

La historia de la cartografía del cerebro humano es, probablemente, más corta de lo que se podría pensar. El avance de los estudios en este campo está asociado con la aparición de herramientas tecnológicas de exploración del cerebro, inicialmente mediante transmisores y tomografías por emisión de positrones (PET). Estas técnicas se usaron en los años ochenta (Herscovitch, 1983) y la primera activación de mapas apareció poco después (Lauter, 1985). Hasta ese momento, las diferencias regionales del cerebro habían sido caracterizadas mediante dibujos hechos a mano por los diferentes investigadores, reduciendo cientos de miles de unidades o ubicaciones en el espacio (voxels) a un pequeño número de mediciones de interés que resultaban válidas anatómicamente, pero muy poco precisas (Ashburner and Friston, 1997).

La idea de utilizar ubicaciones específicas para realizar inferencias estadísticas, a través del uso de Mapas de Modelos Paramétricos (SPM), surgió en respuesta a la necesidad de hacer inferencias sobre las respuestas del cerebro aún desconociendo en que lugar serían expresadas. El primer mapa se utilizó para establecer la especialización funcional que tiene el cerebro para transformar el color (Lueck and Frackowiak, 1989). La metodología utilizada fue descrita en un documento titulado: “The relationship between global and local changes in PET scans”. Esta fue implementada en una gran cantidad de estudios y los procedimientos estadísticos aplicados en la generalidad de ellos fueron análisis de varianza (ANOVA) y análisis de covarianza (ANCOVA) y utilizando como respuesta los promedios de las mediciones en las diferentes regiones de interés como variable respuesta y cada una de las regiones como niveles de un único factor (Lueck and Frackowiak, 1989).

Este enfoque ya se había utilizado en el análisis de los datos de autorradiografías en la neurociencia básica y en el desarrollo metabólico de pacientes y fue también empleado pos-

teriormente con datos de resonancia magnética funcional (fMRI). Pero, se observó que un efecto principal de tratamiento no era suficiente para inferir una respuesta regional específica, debido a que algunos niveles (áreas de interés) inducían un efecto global que dependía en gran medida de la ubicación de dicha unidad (Harmony and Galán, 2001), es decir, no se tenía en cuenta el hecho de que una respuesta medida en una parte del cerebro puede diferir de la respuesta de otra, incluso si la activación neuronal es exactamente igual. Así la convención de usar región como nivel en el análisis no era sostenible (Duffy and Burchfiel, 1981).

Evidentemente, realizar una prueba estadística en cada unidad del espacio (voxel) establecía una enorme tasa de falsos positivos cuando se utilizaban umbrales significativos de decisión sin ser ajustados (Worsley and Evans, 1997). El problema se agravaba por el hecho de que estos datos no eran independiente espacialmente y por tanto una simple corrección de Bonferroni era inadecuada, así después de algunos ajustes y a través de la teoría de procesos estocásticos se establecieron las correcciones necesarias para los valores p (Worsley and Evans, 1997) y el enfoque univariado a través de los SPM continuó su desarrollo, estableciéndose como el método de construcción y evaluación de procesos estadísticos espaciales para probar hipótesis sobre datos de imágenes cerebrales.

Luego de muchos debates se estableció la noción de un espacio anatómico común utilizando rayos X del cráneo como referencia, pero la cuestión entonces era cómo conseguir las imágenes en el espacio de manera eficiente. Al principio, se trató de identificar puntos de referencia (Friston and Penny, 2006). Este enfoque se cayó casi de inmediato puesto que se basaba en la identificación de marcas y no era reproducible cien por ciento. Después se diseñaron plantillas sin la necesidad de puntos de referencia. Las técnicas para la normalización espacial utilizando las plantillas o modelos se han desarrollado de forma constante para producir imágenes específicas en cada paciente (Friston and Penny, 2006).

1.2. Electroencefalograma

Uno de los mecanismos más utilizados en las últimas décadas dentro de la cartografía del cerebro son los registros del Electroencefalograma (EEG). Estos proporcionan registros visibles de la actividad eléctrica generada por las neuronas del cerebro (Friston and Penny, 2006). Dicha actividad queda descrita en términos de varias variables, cada una de las cuales puede ser considerada en relación a la edad y el estado del paciente (Sánchez, 1992). Para hacer el registro de las señales se usan unos discos pequeños de plata con una cubierta de cloruro de plata (electrodos), que son colocados en los lugares del cerebro a estudiar. Algunos elementos importantes del estudio de los registros EEG son su frecuencia, el voltaje, la morfología y la amplitud.

La frecuencia: Se refiere a la actividad del EEG repetitiva (en Hz), o a la frecuencia equi-

valente de ondas simples aisladas (inversa de la duración). Se suelen considerar tres tipos de actividad según la frecuencia: rítmica (ondas de frecuencia aproximadamente constante), arrítmica (no se encuentran presentes ritmos estables), y disrítmica (actividad característica de ciertos grupos de pacientes) (Sánchez, 1992).

Los voltajes: Se refieren al voltaje promedio o al voltaje pico a pico de la actividad del EEG. Los valores del voltaje dependen de la técnica de registro (Sánchez, 1992).

La morfología: se refiere a la forma de la onda. Está determinada por las frecuencias que se combinan y por sus relaciones de fase y voltaje (Sánchez, 1992).

Un electrodo puesto sobre el cráneo registra la suma de potenciales eléctricos procedentes de millones de neuronas de las zonas próximas que están modeladas en su actividad por conexiones; algunas de esas neuronas cada cierto tiempo, liberan su carga eléctrica de manera casi instantánea. Así pues, cada punto en el trazado del EEG es el resultado de la cooperación de una multitud de neuronas. **La amplitud** del trazado habla del tamaño de la población de neuronas que liberó su carga en este tiempo (Sánchez, 1992).

Para el registro de una señal se necesitan al menos dos electrodos; cada uno mide un potencial. Frecuentemente, el potencial de referencia es el de un electrodo colocado en el lóbulo de la oreja, debido a que es un punto con poca actividad eléctrica. En este caso se dice que se trabaja en el modo unipolar o monopolar. El EEG resulta de la diferencia entre estos dos potenciales y no es otra cosa que la gráfica de variación del voltaje con respecto al tiempo.

Otra forma de obtener EEG es determinar la señal de voltaje entre dos electrodos cualesquiera. Ésta se conoce como modo bipolar y puede ser muy útil en el diagnóstico de diferentes enfermedades como epilepsia (en sus diferentes variedades), tumores cerebrales o enfermedades infecciosas que pueden afectar seriamente al cerebro.

1.3. Potenciales evocados (P300)

Son técnicas que registran las respuestas cerebrales provocadas por estímulos sensitivos (visuales, auditivos o táctiles). Se denominan, según el estímulo que produce la respuesta como: Potenciales evocados visuales (PEV), potenciales evocados auditivos (PEA), potenciales evocados somatosensoriales (PESS) o potenciales evocados endógenos o cognitivos. Para registrar estas ondas hay que dar varios cientos de estímulos, y promediar la respuesta recibida, ya que su amplitud es muy baja y puede haber circunstancias externas que provoquen determinadas respuestas. Los potenciales evocados miden el tiempo que le lleva al cerebro recibir e interpretar mensajes (González, 2009).

En sentido amplio un potencial evocado es una variación característica evidenciada en el electroencefalograma y que es producida por un estímulo motor, sensorial o cognitivo, que constituye un indicador importante en procesos elementales y complejos del cerebro (González, 2009). El objetivo básico de investigación en potenciales evocados es detectar diferencias en las respuestas bioeléctricas del cerebro que estén relacionadas con la producción de procesos comportamentales o psicológicos tales como percepción, motivación o aprendizaje y que representen campos de acción al detectar tempranamente cualquier atipicidad presentada. Dentro de dicha investigación se presentan las siguientes fases:

- **Adquisición de la señal EEG:** Una investigación parte típicamente de un diseño experimental en donde la variable independiente es una característica del estímulo como su intensidad o frecuencia. La presentación de los estímulos y la realización de la tarea se hace en sincronía con el registro del EEG. El número de electrodos a utilizar depende de los objetivos de la investigación y es típicamente mayor cuando se pretende no solamente buscar marcadores fisiológicos de procesos cognitivos, sino hacer inferencias sobre las fuentes neurales de las señales (Bauer and Hesselbrock, 1999). Tras el registro, las señales del EEG son amplificadas. Tanto el control del aparataje experimental como la presentación de los estímulos, registro, conversión digital, filtrado y almacenamiento de las señales se realiza de forma informatizada.
- **Procesamiento y Extracción de la Señal:** Tras el registro del EEG, los datos tienen que ser procesados con el fin de mejorar la calidad de las señales. Un aspecto importante es identificar y eliminar aquéllas que provienen de fuentes extracraneales, tales como la actividad muscular o el movimiento. Estas suelen mostrar una mayor amplitud y frecuencia que las señales en el EEG, por lo que pueden eliminarse o atenuarse mediante filtros digitales. Otras fuentes importantes de artefacto son los movimientos oculares y los parpadeos. Dado que éstos producen actividad eléctrica que será propagada por los diferentes electrodos, su actividad se controla mediante electrodos especiales. Los ensayos donde ocurren parpadeos pueden eliminarse o corregirse mediante una serie de algoritmos propuestos en la literatura (Gratton, 1998).
- Otro paso importante en el procesamiento de los datos es la **secuenciación del registro EEG** de forma continua en un conjunto de ensayos, con la actividad en microvoltios medida desde una línea base antes de la aparición del estímulo (típicamente unos cien ms) hasta uno o más segundos después. Dada la baja fiabilidad de las medidas de EEG, las secuencias pertenecientes a una misma condición experimental son promediadas con el fin de reducir el ruido de forma que la señal producida por el efecto experimental sea detectable.

- **Análisis e Interpretación de Resultados:** Una vez calculado el promedio de las ondas, se obtienen las medidas de interés. Una de las más frecuentemente utilizadas son los picos de amplitudes, definidas como la diferencia entre el voltaje máximo y la línea base preestimar para cada condición experimental. Estos resultados se someten a análisis para verificar algunas hipótesis de interés. En éste análisis la estadística ha proporcionado varias herramientas para detectar diferencias, sin embargo en la mayoría de los casos, los datos analizados corresponden a medidas que resumen todas las observaciones en el tiempo y no se ha tenido en cuenta la correlación innata de este tipo de información.

Dentro de los Potenciales evocados cognitivos el más conocido es el llamado ***P300***, que se obtiene cuando el estímulo es relevante para la tarea que ejecuta el individuo. Este tiene una polaridad positiva, una amplitud máxima en la posición parietal y una latencia cercana o superior a 300 milisegundos que depende de la complejidad del estímulo y su función está vinculada a procesos de evaluación perceptiva que en éste caso implica la discriminación entre palabras con diferente carga emocional, neutras y no palabras (combinación de letras si ningún sentido).

2. Marco teórico

En este capítulo se presenta el marco teórico usado como base en la propuesta metodológica. Se dan algunas definiciones básicas de geoestadística univariada y multivariada, de ADF, de regresión y ANOVA funcional y finalmente de geoestadística funcional.

2.1. Geostatística

El enfoque bajo el cual se desarrolla la metodología de predicción y estimación de la estructura de dependencia espacial, se centra en una rama de la estadística que en los últimos tiempos ha cobrado importancia en muchas áreas del conocimiento: ***La estadística Espacial***. Esta, es la reunión de un conjunto de metodologías apropiadas para el análisis de datos de variables aleatorias observadas en diversos sitios (puntos del espacio o agregaciones espaciales) de una región. En general se consideran tres subáreas dentro de su estudio que se diferencian por las características de continuidad del espacio en el cual tienen aplicación: Geoestadística, datos de área o regionales y patrones puntuales. Para el caso de aplicación considerada en el capítulo 4 y dado que las curvas son medidas en diferentes ubicaciones del cerebro, éste es considerado como un dominio continuo y la metodología de análisis apropiada es la geoestadística (Cressie, 1993).

En el contexto de geoestadística, las ubicaciones provienen de un conjunto D continuo y son seleccionadas a juicio del investigador (D fijo). Es importante resaltar que en geoestadística el propósito esencial es la interpolación y si no hay continuidad espacial pueden hacerse predicciones carentes de sentido. Cabe aclarar que el investigador puede hacer selección de puntos del espacio a conveniencia o puede seleccionar los sitios bajo algún esquema de muestreo probabilístico (Giraldo, 2009).

2.1.1. Geoestadística univariada

A continuación se dan elementos básicos teóricos para realizar análisis geoestadísticos en el contexto univariado.

Una variable medida en el espacio de forma que presente una estructura de correlación, se dice que es una variable regionalizada. De manera más formal se puede definir como un proceso estocástico con dominio contenido en un espacio euclidiano d -dimensional R^d , que

generalmente es 1, 2 ó 3, debido a que las coordenadas están referenciadas en un eje, un plano o el espacio $\{Z(s) : s \in D \subset R^d\}$. En términos prácticos $Z(s)$ puede verse como una medición de una variable aleatoria en un punto s de una región de estudio (s representa las coordenadas, planas o geográficas, y Z la variable en cada una de ellas).

Dados s_1, s_2, \dots, s_n , el vector aleatorio $Z(s) = (Z(s_1), Z(s_2), \dots, Z(s_n))^t$ está definido por su función de distribución acumulada conjunta $F(z_1, z_2, \dots, z_n) = P(Z(s_1) \leq z_1, \dots, Z(s_n) \leq z_n)$. El proceso se dice estacionario si $Z(s) = (Z(s_1), Z(s_2), \dots, Z(s_n))^t$ y $Z(s+h) = (Z(s_1+h), Z(s_2+h), \dots, Z(s_n+h))^t$ tienen la misma función de distribución acumulada conjunta para todo h .

Conocidas las densidades marginales univariadas y bivariadas se pueden establecer los siguientes valores esperados (momentos univariados y bivariados):

- $E(Z(s_i)) = m(s_i)$
- $V(Z(s_i)) = E(Z(s_i) - m(s_i))^2 = \sigma_i^2$
- $C(Z(s_i), Z(s_j)) = E[(Z(s_i) - m(s_i))(Z(s_j) - m(s_j))]$: Función de autocovarianza.
- $\gamma(Z(s_i), Z(s_j)) = \frac{1}{2}E(Z(s_i) - Z(s_j))^2$: Función de semivarianza.

Utilizando los momentos de una variable aleatoria, un proceso es estacionario de segundo orden si:

- $E(Z(s)) = \mu$ para todo $s \in D \subset R^d$.
- $Cov(Z(s_i), Z(s_j)) = E[(Z(s_i) - \mu)(Z(s_j) - \mu)] = C(h)$ para todo $s_i, s_j \in D \subset R^d$, $h = \|s_i - s_j\|$.

La existencia de la covarianza implica que la varianza existe, es finita y no depende de h , es decir $V(Z(s_i)) = C(0) = \sigma^2$. Así mismo la estacionariedad de segundo orden implica la siguiente relación entre la función de semivarianza y la de autocovarianza:

$$\gamma(Z(s+h), Z(s)) = C(0) - C(h) = \sigma^2 - C(h) \quad (2.1)$$

Es intrínsecamente estacionario si el proceso $(Z(s), Z(s+h))$ es estacionario, es decir si se cumple:

- $E(Z(s_i) - Z(s_j)) = 0$ para todo $s_i, s_j \in D \subset R^d$.
- $V(Z(s_i) - Z(s_j)) = E[(Z(s_i) - Z(s_j))^2] = 2\gamma(h)$ para todo $s_i, s_j \in D \subset R^d$, $h = \|s_i - s_j\|$.

$C(h)$ es llamada función de autocovarianza y es definida positiva. La función $\gamma(h)$ se denomina función de semivarianza y cumple la propiedad de ser definida negativa condicionalmente, es decir, al ser aplicada a una matriz de distancias, el resultado es una matriz definida negativa (Cressie, 1993). Otra propiedad importante en los procesos geoestadísticos es la condición de isotropía. Un campo aleatorio se dice isotrópico si $C(h)$ depende únicamente de la distancia h entre dos sitios ($|s_i - s_j|$) y no de su dirección (Stein, 1999).

La estimación del variograma (semivarianza) es hecha comúnmente por el método de momentos de la siguiente forma:

$$2\hat{\gamma}(h) = \frac{\sum_{N(h)} [(Z(s_i) - Z(s_j))]^2}{N(h)},$$

dónde $N(h)$ es el número de puntos separados por una distancia h . Generalmente después de haber obtenido el variograma experimental, se modela la estructura de dependencia espacial ajustando un modelo paramétrico (exponencial, gaussiano, esférico, etc.), mediante mínimos cuadrados ponderados (Isaaks and Srivastava, 1987).

Una vez obtenido éste modelo, el siguiente paso es predecir la variable de interés en un sitio no muestreado s_0 , mediante el uso del predictor **kriging** (Cressie, 1993):

$$\hat{Z}(s_0) = \mu(s_0) + \sum_i \lambda_i (Z(s_i) - \mu(s_0)). \quad (2.2)$$

Los parámetros estimados mediante esta metodología se obtienen resolviendo el siguiente conjunto de ecuaciones:

$$\begin{pmatrix} \gamma_{s_1-s_1} & \cdots & \gamma_{s_1-s_n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{s_n-s_1} & \cdots & \gamma_{s_n-s_n} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma_{s_0-s_1} \\ \vdots \\ \gamma_{s_0-s_n} \\ 1 \end{pmatrix}, \quad (2.3)$$

dónde $\gamma_{s_i-s_j}$ es la función de semivarianza entre $Z(s_i)$ y $Z(s_j)$, μ es el multiplicador de lagrange usado para garantizar la condición de insesgadez $\sum_{i=1}^n \lambda_i = 1$. Se resalta que existen varios tipos de kriging que dependen de las restricciones impuestas sobre los pesos λ_i (Cressie, 1993).

2.1.2. Geoestadística multivariada

Si se cuenta con información de más de una variable en cada sitio, es decir, F procesos estocásticos, los procedimientos anteriormente mencionados son extendidos al caso multivariado. El procedimiento de predicción que se utiliza generalmente se denomina **cokriging**.

Sea $\{Z(s) : s \in D \subset R^d\}$ un vector espacial de F procesos aleatorios $Z_1(s), \dots, Z_F(s)$,

sobre una región D , y se asume que cada uno de estos procesos es estacionario. Entonces se definen el variograma y semivariograma cruzado respectivamente:

$$2\gamma_{l,q}(s_i, s_j) = V(Z_l(s_i) - Z_q(s_j)) \quad (2.4)$$

$$\gamma_{l,q}^t = (\gamma_{l,q}(s_1, s_0), \dots, \gamma_{l,q}(s_n, s_0)) \quad (2.5)$$

$$\Gamma_{l,q} = \begin{pmatrix} \gamma_{l,q}(s_1 - s_1) & \dots & \gamma_{l,q}(s_1 - s_n) \\ \vdots & \ddots & \vdots \\ \gamma_{l,q}(s_n - s_1) & \dots & \gamma_{l,q}(s_n - s_n) \end{pmatrix}.$$

El predictor cokriging de la f -ésima variable en el sitio no muestreado s_0 es determinado por una combinación lineal de las demás variables aleatorias en los n sitios muestreados:

$$\hat{Z}(s_0) = \sum_{i=1}^{n_f} \sum_{j=1}^F \lambda_{i,j} Z_i(s_j) \quad (2.6)$$

Mediante multiplicadores de Lagrange se pueden obtener sistemas de ecuaciones similares a los presentados en el caso de kriging (geoestadística univariada, ecuación (2.3)) para obtener los pesos λ_{ij} (Giraldo, 2009).

2.1.3. Modelo lineal de correogionalización (MLC)

Es importante destacar que la matriz de covarianza espacial $C(h)$ debe cumplir la propiedad de ser definida positiva, esto es garantizado solo si la forma funcional ajustada a los variogramas (simples y cruzados) presenta la condición de ser definida negativa. Sin embargo, ajustar un modelo válido para cada uno de los variogramas, no garantiza que el ajuste global tenga la propiedad de ser definido negativo y por lo tanto al utilizar la equivalencia entre la covarianza y la semivarianza (ecuación (2.1)) no se garantiza dicha condición de la matriz de covarianza (dado que una matriz definida por bloques que garantice que cada bloque sea definido positivo, no necesariamente es definida positiva).

Para garantizar que la matriz de covarianza espacial presente la condición de ser definida positiva, el método para ajustar uno de los modelos teóricos conocidos a las estimaciones de semivarianza previamente establecidas, es el modelo lineal de correogionalización. A continuación se presenta en que consiste dicho método.

Un modelo lineal de correogionalización (MLC) asume que todos los semivariogramas simples y cruzados pueden expresarse como una combinación lineal de los mismos modelos teóricos (Isaaks and Srivastava, 1987). Para el caso de dos variables se obtiene:

$$\gamma_l(h) = \alpha_0 \gamma_0(h) + \dots + \alpha_m \gamma_m(h)$$

$$\gamma_k(h) = \beta_0 \gamma_0(h) + \dots + \beta_m \gamma_m(h)$$

$$\gamma_{l,k}(h) = \delta_0 \gamma_0(h) + \dots + \delta_m \gamma_m(h),$$

dónde $\gamma_l(h)$ y $\gamma_k(h)$ son los semivariogramas simples y $\gamma_{l,k}(h)$ es el semivariograma cruzado, $\gamma_0(h), \dots, \gamma_m(h)$ son los modelos teóricos de semivarianza y α_i, β_i y δ_i $i = 1, 2, \dots, m$ son constantes que garantizan que la matriz de correogionalización $\Gamma(h)$ sea definida negativa.

En términos matriciales se tiene:

$$\Gamma(h) = \begin{pmatrix} \gamma_l(h) & \gamma_{l,k}(h) \\ \gamma_{k,l}(h) & \gamma_k(h) \end{pmatrix} = \sum_{s=0}^m B_s \gamma_s(h)$$

dónde:

$$B_s = \begin{pmatrix} \alpha_s & \delta_s \\ \delta_s & \beta_s \end{pmatrix} \quad \gamma_s(h) = \begin{pmatrix} \gamma_s(h) & 0 \\ 0 & \gamma_s(h) \end{pmatrix}$$

Sin embargo, ésta aproximación es válida para el caso en el que se observan más de dos variables en cada uno de los sitios muestreados y garantiza que el modelo ajustado presente la condición de ser definido negativo. Así, aplicando este método para el ajuste de los variogramas simples y cruzados y aplicando la equivalencia (ecuación (2.1)) se garantiza la condición de que la matriz de covarianza espacial $C(h)$ sea definida positiva.

2.2. Análisis de datos funcionales (ADF)

2.2.1. Definiciones

Los desarrollos tecnológicos han hecho posible que los investigadores de muchas áreas dispongan de grandes volúmenes de información para un mismo individuo. Usualmente éstos datos pueden ser representados a través de curvas o en general de funciones. Los métodos estadísticos usuales como ANOVA, regresión o técnicas multivariadas, entre otros, han sido adaptados a esta situación. De ahí surgió un nuevo campo de estudio en estadística denominado Análisis de datos funcionales (ADF).

En el ADF la unidad básica de información es la función completa, más que un conjunto de valores (Ramsay and Dalzell, 1991), es decir, que un dato funcional se puede establecer como la extensión de medidas repetidas cuando la cantidad de mediciones (M) en un determinado individuo es muy grande.

Ferraty and Vieu (2006) definen una variable aleatoria funcional χ , como una variable aleatoria que toma valores en un espacio de funciones, es decir, un espacio infinito dimensional (espacio funcional). Una observación x de la variable aleatoria χ se denomina dato funcional.

Un conjunto de datos funcionales x_1, x_2, \dots, x_n es la observación de n variables funcionales distribuidas como χ . Un dato funcional $x_i(t), t \in T = [a, b] \subset R$, es representado usualmente como un conjunto finito de pares (t_j, x_{ij}) $t_j \in T, j = 1, 2, \dots, M$, donde M representa la cantidad de puntos en los cuales es observada la variable de interés y $y_{ij} = \chi_i(t_j)$ (si no existe error observacional) o $y_{ij} = \chi_i(t_j) + \varepsilon_j$ (en caso contrario).

Sea $T = [a, b] \subset R$. Normalmente se asume que se tienen elementos del espacio:

$$L_2(T) = \left\{ f \rightarrow R, \text{ tal que } \int_t f(t)^2 dt < \infty \right\},$$

con el fin de asegurar la validez de algunas distancias e integrales definidas posteriormente. En general el número de puntos M , es muy grande y entonces métodos de interpolación o de suavizado no paramétrico son utilizados para representar conjunto discretos de pares como una función real. En éste sentido se podría pensar que el análisis de datos funcionales (ADF) es inherente a la metodología de estimación funcional no paramétrica.

2.2.2. Suavizado de funciones mediante funciones B-splines

Una herramienta no paramétrica de mucha utilidad en el ADF es el suavizado de curvas a través de funciones. El procedimiento consiste en aproximar las funciones del espacio considerado a través de (Ramsay and Silverman, 2005):

$$\chi(t) = \sum_{l=1}^K c_l B_l(t) = \mathbf{c}'\mathbf{B}(\mathbf{t}), \quad (2.7)$$

dónde K es el número de funciones de la base.

Aunque existen varias alternativas (B-splines, Fourier, Wavelets, entre otras), en esta sección solo se presentan los conceptos referentes a la base de B-splines, debido a que ésta es la utilizada en la aplicación. Una revisión detallada puede hacerse en (Ramsay and Silverman, 2005).

Dado un conjunto de L puntos interiores del intervalo $T = [a, b]$, es decir, $a = \tau_1 < \tau_2 < \dots < \tau_L = b$, una spline cúbica es una función S definida sobre T , tal que S es un polinomio cúbico en el intervalo $[\tau_{l-1}, \tau_l]$, $l = 1, 2, 3, \dots, L + 1$ con segunda derivada continua (en particular S, S', S'' continuas en todo punto τ_l). Los puntos τ_l son denominados **nodos** y entre mayor sea el número de éstos la curva presentará mayor flexibilidad (Ramsay and Silverman, 2005).

Una spline cúbica S con nodos $\tau_0 < \tau_1 < \dots < \tau_L$, se expresa como:

$$S(t) = \sum_{k=0}^{L+4} c_k B_k(t) = \mathbf{c}'\mathbf{B}(\mathbf{t}), \quad (2.8)$$

donde c es un vector k -dimensional de coeficientes y $B(t)$ es una función $L + 4$ -dimensional con componentes $B_k(t)$.

El método utilizado para aproximar la función χ mediante una base de funciones B-splines se basa en minimizar el sistema:

$$\min_c \sum_{j=1}^M (y_j - S(t_j))^2 + \eta \int_t (S''(t)dt), \quad (2.9)$$

con η un parámetro de penalización para disminuir la variabilidad del ajuste. El número de funciones base (K) y η son estimados a través de procedimientos de validación cruzada (Ramsay and Silverman, 2005). A partir de observaciones (t_j, x_{ij}) $t_j \in T, j = 1, 2, \dots, M$, se pueden obtener n curvas (n datos funcionales) empleando las ecuaciones (2.8) y (2.9).

2.2.3. Herramientas descriptivas del Análisis de datos Funcionales

Dado un conjunto de datos funcionales x_1, x_2, \dots, x_n definidos en $t \subset T \in R$, las correspondientes funciones descriptivas están dadas por las expresiones (Ramsay and Silverman, 2005):

- Media: $\bar{\chi}(t) = \frac{\sum_{i=1}^n \chi_i(t)}{n}$
- Varianza: $Var(\chi(t)) = \frac{\sum_{i=1}^n (\chi_i(t) - \bar{\chi}(t))^2}{n-1}$
- Desviación estándar: $\sqrt{Var(\chi(t))}$
- Covarianza: $Cov(\chi(t_1), \chi(t_2)) = \frac{\sum_{i=1}^n (\chi_i(t_1) - \bar{\chi}(t_1))(\chi_i(t_2) - \bar{\chi}(t_2))}{n-1}$
- Correlación: $Cor(\chi(t_1), \chi(t_2)) = \frac{Cov(\chi(t_1), \chi(t_2))}{\sqrt{Var(\chi(t_1))Var(\chi(t_2))}}$

Así, se puede concluir que las estadísticas descriptivas univariadas y bivariadas clásicas se aplican igualmente cuando se tienen datos funcionales. Sin embargo se resalta que en este caso, los objetos calculados corresponden a curvas.

2.3. Modelos lineales funcionales (Regresión Funcional)

Un modelo lineal permite establecer la relación entre una variable respuesta y una o varias covariables. La extensión al caso funcional se presenta cuando existe por lo menos una variable (respuesta o explicativa) funcional en dicho modelo. Así, el modelo de regresión funcional es la extensión natural del modelo de regresión usual al caso en el cual se cuenta con una variable respuesta funcional y/o con covariables funcionales.

2.3.1. Propuesta de Ramsay and Silverman (2005)

Asumiendo que se tienen G tratamientos cada uno con un número n_g de sujetos y que $y(t)$ es una respuesta funcional, el modelo lineal para la i -ésima función (curva respuesta) en el g -ésimo grupo ($y_{ig}(t)$), está dado por (Ramsay and Silverman, 2005):

$$y_{ig}(t) = \mu(t) + \alpha_g(t) + \varepsilon_{ig}(t), \quad (2.10)$$

dónde la función $\mu(t)$ es la media general, $\alpha_g(t)$ representa la función media para cada tratamiento y $\varepsilon_{ig}(t)$ es la función de error en cada caso. La tarea para determinar el modelo, es establecer la matriz diseño con el fin de estimar los parámetros funcionales $\mu(t)$ y $\alpha_g(t)$, bajo la condición de que $\sum_g \alpha_g(t) = 0$ para todo t con el fin de garantizar la estimabilidad de los mismos.

En términos matriciales el modelo puede expresarse como:

$$y_{ig}(t) = \sum_{j=1}^{(G+1)} x_{(ig)j} \beta_j(t) + \varepsilon_{(ig)}(t) \quad (2.11)$$

$$Y(t) = X(t)\beta(t) + \varepsilon(t) \quad (2.12)$$

Cabe destacar que la matriz diseño X tiene la misma estructura que en el caso de regresión múltiple o simple. La diferencia obedece a que el vector de parámetros $\beta(t)$ y las predicciones $X\beta(t)$, son vectores de funciones en vez de vectores de números.

La estimación de los parámetros funcionales de la ecuación (2.12) por el criterio de mínimos cuadrados, se obtiene minimizando:

$$SSE(\beta) = \sum_g \sum_i^{n_g} \int (y_{ig}(t) - \sum_j^{G+1} x_{(ig)j} \beta_j(t))^2 dt$$

$$SSE(\beta) = \sum_i^n \|y_i(t) - x_i^t \beta(t)\|^2, \quad (2.13)$$

sujeto a la condición de $\sum_{j=2}^{G+1} \beta_j(t) = 0$ para garantizar la estimabilidad de los parámetros. Se destaca que la norma utilizada es la norma euclídea.

Puede observarse que la ecuación (2.13) es una extensión natural de la suma de cuadrados del error (SSE) del modelo lineal clásico (Ramsay and Silverman, 2005).

Básicamente, la mayoría de los estadísticos del análisis de varianza univariado son aplicables al problema funcional, esencialmente, porque el problema de análisis de varianza funcional,

se puede considerar como un análisis de varianza univariado (ANOVA) para cada valor específico t del dominio de la función.

Al igual que en el modelo lineal múltiple, la fuente primaria de información para investigar la importancia de los tratamientos, es la función de suma de cuadrados residual:

$$SSE(t) = \sum_{ig} (y_{ig}(t) - X_{ig}\hat{\beta}(t))^2$$

Dicha función es comparada con la función de suma de cuadrados residual obtenida al utilizar sólo la media general $\hat{\mu}(t)$ en el modelo:

$$SSY(t) = \sum_{ig} (y_{ig}(t) - \hat{\mu}(t))^2$$

Así un camino para realizar la comparación, se establece utilizando la función del cuadrado de correlación múltiple:

$$R^2(t) = \frac{(SSY(t) - SSE(t))}{SSY(t)}$$

o calcular las funciones análogas a las de un ANOVA en el caso univariado. Por ejemplo la función del cuadrado medio del error:

$$CME(t) = \frac{SSE(t)}{gl} = \frac{SSE(t)}{n - G},$$

con gl los grados de libertad asociados en cada caso.

Análogamente, la función de cuadrado medio de la regresión es la diferencia entre la suma de cuadrados total ($SSY(t)$) y la suma de cuadrados del error ($SSE(t)$), dividida por la diferencia entre los grados de libertad del error para los dos modelos (grados de libertad de la regresión):

$$CMR(t) = \frac{SSY(t) - SSE(t)}{gl}$$

Finalmente se puede construir la función F como:

$$F(t) = \frac{CMR(t)}{CME(t)}$$

Sin embargo, se destaca que bajo este enfoque la prueba F (realizada en cada tiempo t) pierde algunas de sus propiedades, debido a que realizar una prueba F en cada tiempo t con un nivel de significancia determinado no implica el mismo nivel de significancia en una prueba conjunta.

2.3.2. Propuesta de Cuevas, Febrero and Fraiman (2004)

Cuevas, Febrero and Fraiman (2004) propusieron un test ANOVA para datos funcionales aplicable al caso de análisis de varianza con un único factor como tratamiento. Este, se adapta al caso en el que las curvas se consideran independientes y homocedásticas (igualdad de funciones de covarianzas), sin embargo, puede ser extendido al caso heterocedástico (diferentes funciones de covarianzas). A continuación se presenta el método establecido en el artículo de éstos autores.

Suponga que $y_{ig}(t)$, $i = 1, 2, \dots, n_g$, $g = 1, 2, \dots, G$, $t \in [a, b]$, son G muestras de trayectorias cuadrado-integrables independientes, tales que $E(y_g(t)) = \mu_g(t) = \mu(t) + \alpha_g(t)$ y que se quiere probar la hipótesis nula:

$$H_0 : \mu_1(t) = \mu_2(t) = \dots = \mu_G(t) \quad (2.14)$$

En una primera aproximación, para el caso homocedástico, se considera la posibilidad de usar una estadística F análoga a la aplicada en el caso univariado:

$$F_n = \frac{\sum_{g=1}^G n_g \|\bar{y}_g - \bar{y}\|^2 / (G - 1)}{\sum_{ig} \|y_{ig} - \bar{y}_g\|^2 / (n - G)}, \quad (2.15)$$

dónde $\bar{y}_g = \sum_{i=1}^{n_g} \frac{y_{ig}(t)}{n_g}$, $\bar{y} = \sum_{g=1}^G \frac{n_g \bar{y}_g(t)}{n}$, $n = \sum_{g=1}^G n_g$ y se utiliza la norma usual de los espacios cuadrado-integrables L^2 :

$$\|y\| = \left(\int_a^b y^2(t) dt \right)^{\frac{1}{2}}$$

Bajo ésta definición, puede verse que mientras el numerador tiene en cuenta la variabilidad “externa” entre las diferentes muestras, el denominador mide la variabilidad “interna” dentro de cada una de ellas. La expresión coincide exactamente con la prueba F clásica, si las funciones y_{ij} son reemplazadas por datos escalares y la norma por el cuadrado de la variable. Sin embargo, la distribución exacta de F_n bajo H_0 no es fácil de deducir, por lo tanto, los autores consideran una prueba basada en el numerador de éste estadístico:

$$T_n = \sum_{g=1}^G n_g \|\bar{y}_g - \bar{y}\|^2, \quad (2.16)$$

y una medida equivalente para medir variabilidad “entre grupos”:

$$V_n = \sum_{g < j} n_g \|\bar{y}_g - \bar{y}_j\|^2 \quad (2.17)$$

En el trabajo de Cuevas, Febrero and Fraiman (2004), se establece que en diseños balanceados, T_n y V_n difieren únicamente por una constante multiplicativa. Nótese que la estadística

V_n presenta una estructura de U – estadística (Borovskikh, 1996), así, las conclusiones pueden ser establecidas utilizando procedimientos bootstrap (Arconés and Giné, 1992).

Procedimiento: Con el fin de aplicar éste test, se debe seguir el siguiente procedimiento:

1. Calcular la función media para cada grupo o tratamiento:

$$\bar{y}_g = \sum_{i=1}^{n_g} \frac{y_{ig}(t)}{n_g}$$

2. Calcular la función media global:

$$\bar{y} = \sum_{g=1}^G \frac{n_g \bar{y}_g(t)}{n}$$

3. Calcular el valor de la estadística de prueba T_n o V_n .
4. Establecer si el valor de la estadística de prueba cae en la región de rechazo. Esta región es establecida mediante procedimientos de remuestreo (bootstrap) utilizados para aproximar la distribución de la estadística de prueba:
 - Generar w muestras de tamaño n con reemplazo a partir de los datos originales.
 - Calcular para cada una de las muestras el valor del estadístico de prueba.
 - Realizar un histograma de estas realizaciones y ubicar en éste el valor del estadístico calculado con los datos originales. Establecer si éste cae dentro de la región de rechazo (extremo superior de la distribución).

2.3.3. Prueba de Shen and Faraway (2004)

Shen and Faraway (2004), propusieron una prueba tipo F para modelos lineales con respuesta funcional. Esta, es una extensión de la prueba F multivariada al caso en que la cantidad de mediciones crece y puede ser utilizada para la comparación de dos modelos anidados cualesquiera. A continuación se presentan los elementos esenciales de la prueba dada por éstos autores. Todos los apartes de esta sección son tomados de su artículo.

Sin pérdida de generalidad, se considera la comparación de dos modelos lineales ω y Ω , donde $\dim(\Omega) = p$ y $\dim(\omega) = q$, $p > q$. El modelo ω resulta de una restricción lineal de los parámetros de Ω . Así:

$$H_0 : Y(t) = X_1 \alpha_1(t) + \epsilon(t) \text{ versus } H_a : Y(t) = X_1 \alpha_1(t) + X_2 \alpha_2(t) + \epsilon(t) \quad (2.18)$$

Sea

$$\mathbf{F} = \frac{(rss_\omega - rss_\Omega)/(p - q)}{rss_\Omega/(n - p)} \approx \frac{\text{traza}(\hat{\Sigma}^\omega - \hat{\Sigma}^\Omega)/(p - q)}{\text{traza}(\hat{\Sigma}^\Omega)/(n - p)}, \quad (2.19)$$

con $rss = \sum_{i=1}^n \int_t (y_i(t) - \hat{y}_i(t))^2 dt$.

Teorema: Asumiendo que el proceso de ruido $\epsilon_i(t)$ es gaussiano e independiente y con función de covarianza continua $r(s, t)$ sobre un intervalo cerrado τ , entonces bajo el sistema de hipótesis dado en (2.18), la estadística F se distribuye como

$$\frac{\sum_{i=1}^{\infty} r_i \chi_{p-q}^2 / (p - q)}{\sum_{i=1}^{\infty} r_i \chi_{n-p}^2 / (n - p)}$$

dónde r_i es el i -ésimo valor propio ordenado de la función de covarianza $r(s, t)$ y las variables χ^2 son independientes.

En su artículo los autores demuestran que la estadística F tiene una distribución denominada distribución funcional F con coeficientes $r_i, i = 1, 2, \dots, \infty$ y grados de libertad $(p - q, n - p)$. Así mismo, realizan la aproximación de esta distribución a la distribución F de Fisher con grados de libertad f_1 y f_2 :

$$f_1 = \frac{(\sum_{i=1}^{\infty} r_i)^2}{\sum_{i=1}^{\infty} r_i^2} (p - q) \quad (2.20)$$

$$f_2 = \frac{(\sum_{i=1}^{\infty} r_i)^2}{\sum_{i=1}^{\infty} r_i^2} (n - p) \quad (2.21)$$

En la práctica, solo se observa $y_i(t)$ sobre una grilla de puntos $t_j, j = 1, 2, \dots, M$, así, el factor de ajuste $\frac{(\sum_{i=1}^{\infty} r_i)^2}{\sum_{i=1}^{\infty} r_i^2}$ puede ser estimado por $\frac{\text{traza}(E)^2}{\text{traza}(E'E)}$, dónde $E = \hat{\Sigma}^\Omega$ es la matriz de covarianza empírica del ruido calculada con el modelo completo Ω .

Procedimiento: Con el fin de aplicar el test de Shen and Faraway (2004) se debe:

1. Establecer la matriz diseño para el modelo completo (Ω) y para el modelo reducido (ω).
2. Estimar los parámetros funcionales de cada uno de los modelos ($\alpha_1(t)$ y $\alpha_2(t)$ de la ecuación (2.18)).

3. Calcular las matrices de sumas de cuadrados ($\hat{\Sigma}^\Omega$ y $\hat{\Sigma}^\omega$) respectivas mediante:

$$rss = \sum_{i=1}^n \int_t (y_i(t) - \hat{y}_i(t))^2 dt \quad (2.22)$$

$$\hat{\Sigma}^\omega = (y(t) - \hat{y}^\omega(t))'(y(t) - \hat{y}^\omega(t)) \quad (2.23)$$

$$\hat{\Sigma}^\Omega = (y(t) - \hat{y}^\Omega(t))'(y(t) - \hat{y}^\Omega(t)) \quad (2.24)$$

$$\text{Con: } y(t) - \hat{y}(t) = \begin{pmatrix} y_1(t_1) - \hat{y}_1(t_1) & y_1(t_2) - \hat{y}_1(t_2) & \dots & y_1(t_M) - \hat{y}_1(t_M) \\ y_2(t_1) - \hat{y}_2(t_1) & y_2(t_2) - \hat{y}_2(t_2) & \dots & y_2(t_M) - \hat{y}_2(t_M) \\ \vdots & \vdots & \ddots & \vdots \\ y_n(t_1) - \hat{y}_n(t_1) & y_n(t_2) - \hat{y}_n(t_2) & \dots & y_n(t_M) - \hat{y}_n(t_M) \end{pmatrix}$$

Se destaca que las estimación $\hat{y}_i(t_j)$ se establecen utilizando el modelo correspondiente (reducido o completo).

4. Calcular la estadística de prueba F (ecuación (2.19)) y los respectivos grados de libertad f_1 y f_2 (ecuaciones (2.20) y (2.21)).
5. Determinar si el valor de la estadística de prueba cae en la zona de rechazo o calcular el valor p con el fin de tomar una decisión acerca de la hipótesis enunciada en la ecuación (2.18).

2.3.4. Prueba de Cuesta-Albertos and Febrero-Bande (2007)

Cuesta-Albertos and Febrero-Bande (2007), proponen un test ANOVA para datos funcionales basado en proyecciones aleatorias. La prueba está diseñada para determinar si existe efecto de factores principales y de interacción así como para determinar efectos de covariables funcionales en un análisis a dos vías. A continuación se presenta de manera resumida la propuesta de dichos autores.

Sean R y $S \in N$ y asuma que para todo $g = 1, 2, \dots, G$ y $k = 1, 2, \dots, S$ existen y_{igk} $i = 1, 2, \dots, n_{gs} \in N$, funciones aleatorias en un espacio de Hilbert separable, tales que:

$$y_{igk}(t) = \mu(t) + \alpha_g(t) + \beta_k(t) + \rho_{gk}(t) + \gamma(t)\Upsilon_{igk} + \epsilon_{igk}, \quad (2.25)$$

dónde:

- La función $\mu(t)$ es no aleatoria y describe la forma del proceso.

- Las funciones no aleatorias $\alpha_g(t)$, $\beta_k(t)$ y $\rho_{gk}(t)$ pertenecen al espacio de Hilbert y toman en consideración el efecto de los factores principales y de la interacción entre éstos, respectivamente.

Con miras a que el modelo sea identificable, se asume que:

$$\sum_g \alpha_g(t) = \sum_k \beta_k(t) = \sum_g \rho_{g,k_0}(t) = \sum_k \rho_{g_0,k}(t) = 0$$

- Υ_{igk} son cantidades aleatorias conocidas que influyen el proceso a través de las funciones γ . Juegan el mismo rol de covariables en el caso unidimensional.
- Las trayectorias aleatorias ϵ_{igk} se asumen independientes y centradas en media. Para cada g y k fijos, ϵ_{igk} son idénticamente distribuidos.

El interés es probar las hipótesis nulas:

$$H_0^A : \alpha_1(t) = \dots = \alpha_G(t) = 0 \quad (2.26)$$

$$H_0^B : \beta_1(t) = \dots = \beta_S(t) = 0 \quad (2.27)$$

$$H_0^I : \rho_{1,1}(t) = \dots = \rho_{G,S}(t) = 0 \quad (2.28)$$

$$H_0^C : \gamma = 0 \quad (2.29)$$

Teorema: Bajo las condiciones anteriores, sea μ^* una distribución Gaussiana sobre un espacio de Hilbert separable H , tal que cada una de sus proyecciones unidimensionales son no-degeneradas. Entonces:

- Si existe g_1, g_2 tal que $\alpha_{g_1} \neq \alpha_{g_2}$, entonces

$$\mu^* \{v \in H : \text{tal que } \langle v, \alpha_1 \rangle = \dots = \langle v, \alpha_G \rangle\} = 0$$

- Si existe k_1, k_2 tal que $\beta_{k_1} \neq \beta_{k_2}$, entonces

$$\mu^* \{v \in H : \text{tal que } \langle v, \beta_1 \rangle = \dots = \langle v, \beta_S \rangle\} = 0$$

- Si existe $(g_1, k_1), (g_2, k_2)$ tal que $\rho_{(g_1, k_1)} \neq \rho_{(g_2, k_2)}$, entonces

$$\mu^* \{v \in H : \text{tal que } \langle v, \rho_{(1,1)} \rangle = \dots = \langle v, \rho_{(G,S)} \rangle\} = 0$$

- Si $\gamma = 0$, entonces

$$\mu^* \{v \in H : \text{tal que } \langle v, \gamma \rangle\} = 0$$

En su artículo, los autores proponen un procedimiento utilizando el anterior teorema con el fin de probar cualquiera de las hipótesis de interés. A continuación se resume dicho procedimiento:

Sean $Y_g(t)$ variables aleatorias funcionales con realizaciones $y_{ig}(t)$ $i = 1, 2, \dots, n_g$ y $g = 1, 2, \dots, G$.

$Y_1(t)$	$Y_2(t)$	\dots	$Y_G(t)$
$y_{11}(t)$	$y_{12}(t)$	\dots	$y_{1G}(t)$
\vdots	\vdots	\ddots	\vdots
$y_{n1}(t)$	$y_{n2}(t)$	\dots	$y_{nG}(t)$

El interés está en probar la hipótesis nula

$$H_0^f : \alpha_1(t) = \alpha_2(t) = \dots = \alpha_G(t), \quad (2.30)$$

con $\alpha_g(t) = E(Y_g(t))$, $g = 1, 2, \dots, G$

Procedimiento: Si se observa la curva en T tiempos espaciados w unidades:

1. Definir un movimiento browniano $v(t)$ de la siguiente forma:

- Sea Z_1, Z_2, \dots, Z_T variables normales estándar independientes.
- Sea $v(0) = 0$ y para cada $h = 1, 2, \dots, T$ $t_h = w \times h$ y

$$v(t_h) = v(t_{h-1}) + w \times Z_h \quad (2.31)$$

2. Defina la proyección de la curva $y_{ig}(t)$ sobre el espacio generado por v como:

$$y_{ig}^v = \int_t y_{ig}(t)v(t)dt, i = 1, 2, \dots, n; g = 1, 2, \dots, G. \quad (2.32)$$

Y_1^v	Y_2^v	\dots	Y_G^v
$y_{11}^v = \int_t y_{11}(t)v(t)dt$	$y_{12}^v = \int_t y_{12}(t)v(t)dt$	\dots	$y_{1G}^v = \int_t y_{1G}(t)v(t)dt$
\vdots	\vdots	\ddots	\vdots
$y_{n1}^v = \int_t y_{n1}(t)v(t)dt$	$y_{n2}^v = \int_t y_{n2}(t)v(t)dt$	\dots	$y_{nG}^v = \int_t y_{nG}(t)v(t)dt$

Así, el modelo (ecuación (2.25)) se transforma en:

$$y_{ig}^v = \mu^v + \alpha_g^v + \beta_k^v + \rho_{gk}^v + \gamma \Upsilon_{igk}^v + \epsilon_{igk}^v \quad (2.33)$$

El interés está en probar la hipótesis:

$$H_0^v : \alpha_1^v = \alpha_2^v = \dots = \alpha_G^v \quad (2.34)$$

3. Nótese que los y_{ig}^v , $i = 1, 2, \dots, n$; $g = 1, 2, \dots, G$ son valores de una variable aleatoria y con éstos es posible realizar análisis clásicos como pruebas t o F .
4. Así, si H_0^f (ecuación (2.30)) se cumple, entonces $\alpha_1^v = \alpha_2^v = \dots = \alpha_G^v$, dado que por el teorema mencionado, si $\alpha_g(t) \neq \alpha_r(t)$ entonces es imposible que $\alpha_g^v = \alpha_r^v$ y consecuentemente la prueba es equivalente a probar directamente la hipótesis H_0^f (ecuación (2.30)) o la hipótesis nula (condicional sobre v , ecuación (2.34)) H_0^v . α_g^v denota la media del g -ésimo tratamiento $E(Y_g^v)$ calculada a partir de las observaciones y_{ig}^v . Los autores demuestran que probar H_0^f es equivalente a probar H_0^v .
5. Se realiza un análisis de varianza univariado en el que la respuesta es la proyección de cada una de las curvas sobre el espacio generado por el movimiento browniano y_{ig}^v y la matriz diseño está determinada por los respectivos tratamientos.

2.4. Geoestadística de datos funcionales

Para el caso de la aplicación considerada en el capítulo 4, las herramientas utilizadas corresponden a geoestadística de datos funcionales. A continuación se precisan los conceptos básicos de éstas herramientas.

Sea $\{\chi_s(t) : s \in D \subseteq R^d, t \in T = [a, b] \subset R\}$ un proceso aleatorio tal que $\chi_s(t)$ es una variable aleatoria funcional. Sean s_1, s_2, \dots, s_n puntos en D , en los cuales se observa una realización del proceso aleatorio $\{\chi_s(t) : \chi_{s_1}(t), \chi_{s_2}(t), \dots, \chi_{s_n}(t)\}$.

El proceso es estacionario de segundo orden e isotrópico si:

- $E(\chi_s(t)) = m(t)$, para todo $t \in T$; $s \in D$.
- $V(\chi_s(t)) = \sigma^2(t)$, para todo $t \in T$; $s \in D$.
- $Cov(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h; t) = C_{s_i s_j}(t)$, para todo $s_i, s_j \in D$; $t \in T$; donde $h = \|s_i - s_j\|$.
- $\frac{1}{2}V(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t) = \gamma_{s_i s_j}(t)$; para todo $s_i, s_j \in D$; $t \in T$; donde $h = \|s_i - s_j\|$.

$\gamma(h; t)$ es llamada semivariograma. El objetivo central al igual que en el caso univariado o multivariado es predecir el valor de la variable en un lugar no muestreado s_0 , en este caso

una variable funcional.

El predictor kriging ordinario adaptado al contexto funcional es definido como (Giraldo, 2009):

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi_{s_i}(t), \quad \lambda_1, \dots, \lambda_n \in R. \quad (2.35)$$

Este predictor dado en la ecuación (2.35) tiene la misma expresión del kriging ordinario, pero considerando curvas en vez de funciones. La curva predicha es obtenida como combinación lineal de curvas observadas. Este mismo fue considerado inicialmente por Goulard and Voltz (1991). Los pesos λ_i óptimos para la predicción se obtienen al solucionar el siguiente conjunto de ecuaciones (Giraldo, 2009):

$$\begin{pmatrix} \int_t \gamma_{s_1, s_1}(t) dt & \dots & \int_t \gamma_{s_1, s_n}(t) dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_t \gamma_{s_n, s_1}(t) dt & \dots & \int_t \gamma_{s_n, s_n}(t) dt & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \int_t \gamma_{s_0, s_1}(t) dt \\ \vdots \\ \int_t \gamma_{s_0, s_n}(t) dt \\ 1 \end{pmatrix}, \quad (2.36)$$

La función $\gamma(h) = \int_t \gamma_{s_i s_j}(t) dt = \frac{1}{2} \int_t V(\chi_{s_i}(t) - \chi_{s_j}(t)) dt$, $h = \|s_i - s_j\|$, se denomina traza-variograma, y corresponde a la extensión del semivariograma al caso en el que las mediciones son curvas (Giraldo, 2009).

Usando el teorema de Fubini la traza-variograma se puede expresar como:

$$\gamma(h) = \frac{1}{2} \int_t E((\chi_{s_i}(t) - \chi_{s_j}(t))^2) dt = \frac{1}{2} E \left[\int_t (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \right] \quad (2.37)$$

y por consiguiente su estimador de momentos es:

$$\hat{\gamma}(h) = \frac{\sum_{i,j \in N(h)} \int_t [\chi_{s_i}(t) - \chi_{s_j}(t)]^2 dt}{2N(h)}, \quad (2.38)$$

dónde $N(h)$ es el número de parejas de sitios separados por una distancia h ($h = \|s_i - s_j\|$).

Una vez estimada la traza variograma para la variable de interés y para una secuencia de I valores h_i , se propone ajustar un modelo paramétrico (por ejemplo el exponencial, gaussiano, esférico), a los puntos $(h_i, \hat{\gamma}(h_i))$ para obtener $\int_t \gamma_{s_i, s_j}(t) dt$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$

La covarianza integrada se obtiene a través de la equivalencia (Giraldo, 2009):

$$\int_T C_{l,q}(s_i, s_j)(t) = \int_T \sigma_{l,q}^2(t) - \int_T \gamma_{l,q}(s_i, s_j)(t). \quad (2.39)$$

3. ANOVA funcional espacialmente correlacionado

En este capítulo se presenta una propuesta metodológica para realizar análisis de varianza de datos funcionales espacialmente correlacionados. Se realizan extensiones de los modelos ANOVA del capítulo anterior. Se propone usar geoestadística funcional para estimar la estructura de dependencia espacial entre y dentro de los tratamientos. Posteriormente ésta es empleada en la estimación por mínimos cuadrados generalizados (MCG) de los parámetros y en la realización de pruebas de hipótesis.

3.1. Transformación de los datos a datos funcionales

Suponga que se tienen datos funcionales $x_i(t), t \in T = [a, b] \subset R$, en cada uno de n sitios de un dominio continuo bajo igualdad de condiciones experimentales (tratamientos). El objetivo es establecer si hay diferencias en las respuestas respecto a los tratamientos. Usualmente $x_i(t)$ es observado en un conjunto finito de puntos t_j y por lo tanto se dispone de pares (t_j, x_{ij}) $t_j \in T, j = 1, 2, \dots, M$. En la primera etapa del análisis se propone usar los métodos descritos en la sección 2.2.2 para obtener curvas a partir del conjunto discreto de pares.

En éste trabajo, se emplean bases de funciones B-splines y validación cruzada no paramétrica para determinar el número óptimo de funciones de la base (K) y el valor del parámetro de penalización η (ver sección 2.2.2). No obstante las otras alternativas de suavizado son igualmente válidas.

3.2. Estimación de la estructura de dependencia espacial

Con los procedimientos de geoestadística funcional dados en la sección 2.4 se propone estimar la estructura de correlación espacial. Se resalta que dicha estimación comprende involucrar la dependencia espacial dentro de cada tratamiento, es decir, entre las curvas medidas dentro de una misma condición experimental. Y entre tratamientos, es decir, entre las curvas medidas

en condiciones experimentales diferentes. A continuación se presentan las etapas necesarias para obtener la matriz de covarianza espacial:

1. Estimar la estructura de correlación espacial dentro de cada tratamiento:

Para la primera estructura de dependencia espacial (entre las curvas medidas dentro de una misma condición experimental), se debe estimar la función traza-variograma descrita en la sección 2.4 (ecuación (2.38)).

2. Estimar la estructura de correlación espacial entre dos tratamientos:

Para la segunda estructura de dependencia espacial (entre las curvas medidas en dos condiciones experimentales diferentes), se propone extender el concepto de traza variograma (ecuación (2.37)) al concepto de traza variograma cruzada mediante:

$$\gamma(h) = \frac{1}{2} \int_t E((\chi_{s_i}^l(t) - \chi_{s_j}^j(t))^2) dt = \gamma(h) = \frac{1}{2} E \left[\int_t (\chi_{s_i}^l(t) - \chi_{s_j}^j(t))^2 dt \right],$$

y por consiguiente su estimador de momentos (MoM) es:

$$2\hat{\gamma}_c(h) = \frac{\sum_{i,j \in N(h)} \int_T [\chi_{s_i}^l(t) - \chi_{s_j}^j(t)]^2 dt}{N(h)} \quad (3.1)$$

3. Ajuste de un modelo paramétrico a las traza variogramas simples y cruzados estimados en 1 y 2:

Una vez son obtenidas las estimaciones de las traza variogramas simples (para cada uno de los tratamientos de interés), y las de las traza variogramas cruzadas (para cada par de tratamientos), para una secuencia de I valores h_i , se propone ajustar un modelo paramétrico de los conocidos en geoestadística univariada (exponencial, gaussiano, Mattern, esférico, etc.) para obtener la estructura de dependencia espacial a cualquier posible distancia h . Finalmente, si $\hat{\gamma}(h)$ denota la estimación paramétrica de las traza variogramas simples y cruzadas, esta forma funcional es utilizada para estimar la estructura de correlación espacial a través de la equivalencia entre el variograma y la covarianza en procesos estacionarios. En geoestadística univariada se tiene la relación (Cressie, 1993) :

$$C(h) = \sigma^2 - \gamma(h).$$

Esta es extendida al contexto funcional mediante (Giraldo, 2009):

$$\int_t C(h; t) = \int_t \sigma^2(t) - \int_t \gamma(h; t) \quad (3.2)$$

La matriz de covarianza espacial debe cumplir la propiedad de ser definida positiva. Esto se garantiza solo si la forma funcional utilizada para ajustar las traza variogramas presenta la condición de ser definida negativa (Cressie, 1993). Sin embargo, ajustar un modelo válido para cada una de las traza variogramas simples y cruzadas, no garantiza que el ajuste global tenga la propiedad de ser definido negativo y por lo tanto al utilizar la equivalencia anterior no se garantiza la condición de definición positiva de la matriz de covarianza (dado que una matriz definida por bloques que garantice que cada bloque sea definido positivo, no necesariamente es definida positiva). Para obtener una matriz de covarianza válida puede usarse un modelo lineal de correogionalización (Isaaks and Srivastava, 1987).

3.3. Estimación de los parámetros del ANOVA funcional. Extensión de la propuesta de Ramsay and Silverman (2005)

Aquí se presenta una aproximación extendiendo la metodología de regresión funcional múltiple (Ramsay and Silverman, 2005) mediante mínimos cuadrados generalizados.

Se propone emplear mínimos cuadrados generalizados usando la matriz de covarianza estimada en la sección 3.2. El modelo lineal de interés es:

$$y(t) = X\beta(t) + \epsilon(t),$$

Por el criterio de mínimos cuadrados generalizados se deben encontrar los $\beta(t)$ que minimizan:

$$GSSE(\beta) = \int_t [y(t) - X\beta(t)]' \Omega^{-1} [y(t) - X\beta(t)] dt, \quad (3.3)$$

dónde la matriz Ω está definida como la matriz de varianza y covarianza del proceso de ruido $\epsilon(t)$ condicionada a la información proporcionada por X .

Asumiendo que tanto las funciones observadas $y_i(t)$ y los parámetros funcionales de regresión $\beta_j(t)$ son expresados a través de funciones base se tiene:

$$y(t) = A\phi(t)$$

$$\begin{pmatrix} \hat{y}_1(t) \\ \hat{y}_2(t) \\ \vdots \\ \hat{y}_n(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_k(t) \end{pmatrix}$$

$$\beta(t) = B\theta(t)$$

$$\begin{pmatrix} \hat{\beta}_1(t) \\ \hat{\beta}_2(t) \\ \vdots \\ \hat{\beta}_p(t) \end{pmatrix} = \begin{pmatrix} \hat{b}_{11} & \hat{b}_{12} & \cdots & \hat{b}_{1k'} \\ \hat{b}_{21} & \hat{b}_{22} & \cdots & \hat{b}_{2k'} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{b}_{p1} & \hat{b}_{p2} & \cdots & \hat{b}_{pk'} \end{pmatrix} \begin{pmatrix} \theta_1(t) \\ \theta_2(t) \\ \vdots \\ \theta_{k'}(t) \end{pmatrix}$$

Con esta reparametrización, la función a minimizar es:

$$GSSE(\beta) = \int_t (A\phi(t) - XB\theta(t))' \Omega^{-1} (A\phi(t) - XB\theta(t)) dt \quad (3.4)$$

$$= \text{traza}(A'\Omega^{-1}AJ_{\phi\phi}) - 2\text{traza}(BJ_{\theta\theta}A'\Omega^{-1}X) + \text{traza}(X'\Omega^{-1}XBJ_{\theta\theta}B')$$

Derivando en la ecuación (3.4) con respecto a B e igualando a cero, se obtiene la solución por mínimos cuadrados generalizados:

$$Z'\Omega^{-1}ZBJ_{\theta\theta} = X'\Omega^{-1}AJ_{\phi\theta} \quad (3.5)$$

$$\hat{B} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}AJ_{\phi\theta}J_{\theta\theta}^{-1} \quad (3.6)$$

$$\hat{\beta} = \hat{B}\theta \quad (3.7)$$

3.4. Extensión de la prueba de Shen and Faraway (2004)

La prueba de Shen and Faraway (2004), está basada en el supuesto de independencia, debido a que en el cálculo de la suma de cuadrados de los residuales no interviene la estructura de correlación espacial. Los residuales definidos y con los cuales se calculan las respectivas sumas de cuadrados suponen que las curvas son independientes. En la versión sin correlación de éste test las matrices de sumas de cuadrados residuales son calculadas mediante las ecuaciones (2.23) y (2.24) de la sección 2.3.3.

Debido a que en varias áreas de interés, el supuesto de independencia es inválido (las curvas presentan diferentes estructuras de correlación), la matriz de varianzas y covarianzas es diferente a la matriz identidad y ésta debe ser incluida en el proceso de estimación de los parámetros y en la obtención de las sumas de cuadrados de los residuales, con el fin de garantizar todos los supuestos del teorema de Gauss-Markov. En particular el supuesto de que los

estimadores β resulten de mínima varianza (Draper and Smith, 1981). Siguiendo la conocida extensión de mínimos cuadrados ordinarios a mínimos cuadrados generalizados (Draper and Smith, 1981), las sumas de cuadrados se deben calcular ponderando por la inversa de la matriz de covarianza espacial del proceso. Esta estructura es estimada a través del modelo paramétrico ajustado a las traza variogramas (simples y cruzados) mediante el modelo lineal de correogionalización. Así, bajo el enfoque de correlación espacial:

$$rss_{\Omega} = \sum_{i=1}^n \int_t (y_i(t) - \hat{y}_i^{\Omega}(t))' \Omega^{-1} (y_i(t) - \hat{y}_i^{\Omega}(t)) dt \quad (3.8)$$

$$rss_{\omega} = \sum_{i=1}^n \int_t (y_i(t) - \hat{y}_i^{\omega}(t))' \Omega^{-1} (y_i(t) - \hat{y}_i^{\omega}(t)) dt \quad (3.9)$$

$$\hat{\sum}^{\omega} = (y(t) - \hat{y}^{\omega}(t))' \Omega^{-1} (y(t) - \hat{y}^{\omega}(t)) \quad (3.10)$$

$$\hat{\sum}^{\Omega} = (y(t) - \hat{y}^{\Omega}(t))' \Omega^{-1} (y(t) - \hat{y}^{\Omega}(t)) \quad (3.11)$$

dónde:

$$y(t) - \hat{y}(t) = \begin{pmatrix} y_1(t_1) - \hat{y}_1(t_1) & y_1(t_2) - \hat{y}_1(t_2) & \dots & y_1(t_M) - \hat{y}_1(t_M) \\ y_2(t_1) - \hat{y}_2(t_1) & y_2(t_2) - \hat{y}_2(t_2) & \dots & y_2(t_M) - \hat{y}_2(t_M) \\ \vdots & \vdots & \ddots & \vdots \\ y_n(t_1) - \hat{y}_n(t_1) & y_n(t_2) - \hat{y}_n(t_2) & \dots & y_n(t_M) - \hat{y}_n(t_M) \end{pmatrix}$$

Se destaca que estos residuales y las sumas de cuadrados presentan las mismas propiedades distribucionales asumidas por los autores, lo que valida la aplicación de dicha prueba.

Una vez obtenidas las matrices de sumas de cuadrados de los residuales a través de la ecuaciones (3.10) y (3.11) (tanto del modelo completo Ω como del reducido ω), se realiza el procedimiento para probar la hipótesis descrita en la ecuación (2.18).

Procedimiento: A continuación se presentan las etapas para realizar dicha prueba:

1. Establecer la matriz diseño para el modelo completo (Ω) y para el modelo reducido (ω) presentados en la ecuación (2.18).
2. Estimar los parámetros funcionales de cada uno de los modelos ($\alpha_1(t)$ y $\alpha_2(t)$) de la ecuación (2.18), involucrando la estructura de dependencia espacial mediante mínimos cuadrados generalizados a través de la ecuaciones (3.5), (3.6) y (3.7).
3. Calcular las matrices de sumas de cuadrados respectivas ($\hat{\sum}^{\Omega}$ y $\hat{\sum}^{\omega}$) descritas en las ecuaciones (3.10) y (3.11).

4. Calcular la estadística de prueba F de la ecuación (2.19) y los respectivos grados de libertad f_1 y f_2 de las ecuaciones (2.20) y (2.21).
5. Determinar si el valor de la estadística de prueba cae en la zona de rechazo (acudiendo a la distribución F de Fisher) o calcular el valor p con el fin de tomar una decisión acerca de la hipótesis enunciada en la ecuación (2.18).

3.5. Extensión de la Prueba de Cuesta-Albertos and Febrero-Bande (2007)

Para involucrar la estructura de correlación espacial en la prueba propuesta por éstos autores, se realiza un ANOVA unidimensional correlacionado espacialmente. En éste, la variable respuesta es la proyección de cada una de las curvas sobre el espacio generado por el movimiento browniano definido en la ecuación (2.31) (nótese que ésta respuesta es escalar). Para estimar la estructura de correlación espacial se utilizan los métodos descritos en la sección 2.1. Para realizar la prueba ANOVA, se usa la estadística F usual con errores correlacionados. A continuación se presentan las etapas necesaria para la aplicación de la prueba de éstos autores (Cuesta-Albertos and Febrero-Bande, 2007) en el contexto de dependencia espacial.

Si se cuenta con un solo factor con diferentes niveles como tratamientos, la ecuación (2.25) es transformada en:

$$y_{ig}(t) = \mu(t) + \alpha_g(t) + \epsilon_{ig}(t) \quad (3.12)$$

El interés está en probar la hipótesis descrita en la ecuación (2.30) con $\alpha_g(t) = E(Y_g(t))$, $g = 1, 2, \dots, G$.

Procedimiento para realizar la prueba propuesta por (Cuesta-Albertos and Febrero-Bande, 2007)

1. Dadas $Y_g(t)$ variables aleatorias funcionales con realizaciones $y_{ig}(t)$ $i = 1, 2, \dots, n$ y $g = 1, 2, \dots, G$, encontrar la proyección ortogonal y_{ig}^v , $i = 1, 2, \dots, n$; $g = 1, 2, \dots, G$ (ecuación (2.32)) de cada una de éstas sobre el espacio generado por el movimiento browniano descrito en la ecuación (2.31).
2. Una vez obtenidas y_{ig}^v , éstas son consideradas como variable respuesta en un modelo de análisis de varianza univariado usual con correlación espacial. Los tratamientos están determinados por los mismos niveles del factor considerado en el caso funcional $\alpha_g(t)$ (ecuación (3.12)).
3. En este análisis de varianza, se involucra la estructura de dependencia espacial obtenida a través del ajuste de un modelo paramétrico a la estimación de los semivariogramas simples y cruzados (sección 2.1) de los tratamientos mediante MLC.

4. La estimación de los parámetros escalares α_g^v (ecuación (2.33)) es realizada mediante mínimos cuadrados generalizados.
5. Una vez obtenidos los parámetros estimados en el ANOVA, se procede a calcular las sumas de cuadrados residuales involucrando dicha estructura de dependencia espacial.
6. Con éstas sumas de cuadrados residuales, se lleva a cabo una prueba F clásica.

3.6. Resumen

Como un resumen general, se muestra un cuadro donde se integran los procedimientos vistos y los requisitos mínimos para su aplicabilidad.

Etapa/Caso	Sin correlación (Versión clásica)	Con correlación (Extensiones propuestas)
Estimación de los parámetros funcionales	Mínimos cuadrados ordinarios: Propuesta de Ramsay and Silverman (2005), ecuación (2.13)	Mínimos cuadrados generalizados: Propuesta realizada en éste trabajo, ecuación (3.8).
Prueba de Cuevas, Febrero and Fraiman (2004)	Solo es aplicable a análisis de varianza con un factor de clasificación, ecuaciones (2.16) y (2.17).	No tiene aplicación, dado que se necesita bootstrap para su implementación.
Prueba de Shen and Faraway (2004)	Las sumas de cuadrados se calculan con las ecuaciones (2.23) y (2.24).	Se debe estimar la matriz de correlación espacial a través de los procedimientos de geoestadística funcional descritos en la sección 2.4 y en la sección 3.2. Posteriormente incluir dicha matriz en el cálculo de las sumas de cuadrados de los residuales a través de los propuesto en éste trabajo en las ecuaciones (3.11) y (3.12).
Prueba de Cuesta-Albertos and Febrero-Bande (2007)	La prueba se realiza con un análisis de varianza clásico en el cual la variable respuesta es la proyección ortogonal de cada curva sobre el espacio generado por el movimiento Browniano.	Se debe realizar un análisis de varianza escalar con errores correlacionados espacialmente. El procedimiento para realizar dicho análisis se propone en la sección 3.5. La matriz de covarianza se estima a través de procedimientos de geoestadística multivariada.

Tabla 3.1.: Resumen de aplicabilidad de las pruebas clásicas y de las pruebas propuestas

4. Aplicación

En éste capítulo se presenta la aplicación de las metodologías propuestas en el capítulo 3 a datos de curvas cerebrales que corresponden a electroencefalogramas registrados como respuesta al potencial evocado P300.

En la primera parte se hace una descripción de la base de datos analizada y de los procedimientos aplicados para establecer las coordenadas de los sitios muestreados (ubicaciones de los electrodos). En la segunda se hacen análisis de varianza funcionales en los que es válido el supuesto de independencia entre individuos. En éstos, se usan los procedimientos clásicos. En la última parte se considera el espacio como un factor relevante y se usan por consiguiente las propuestas realizadas en el capítulo 3.

4.1. Descripción de la base de datos

El experimento que dió origen a los datos considerados en ésta aplicación, se basa en que a cada paciente se le mostraba un conjunto de palabras con carga emocional positiva como abrazo, paz y regalo, negativa como guerra, enfermedad y muerte o neutra como número y roca y también una serie de no palabras (combinaciones de letras sin ningún sentido lógico). La tarea de cada uno de ellos consistía en responder mediante un botón adyacente cuando aparecía únicamente una palabra. El número de pacientes total en el experimento fue 48; 30 hombres y 18 mujeres.

La presentación de los estímulos y la realización de la tarea se hizo en sincronía con el registro del EEG, con lo cual se obtuvo la medición de la actividad del cerebro registrada como la diferencia entre el voltaje máximo y la línea base preestímulo (electrodo determinado ubicado cerca a la oreja). Esto se hizo para cada condición experimental. En cada paciente se obtuvo la información detallada en 19 electrodos dispuestos en ubicaciones establecidas por el sistema 10-20.

4.1.1. Sistema de referencia 10-20

El sistema Internacional de referencia 10-20 (Hasper, 1958) fue diseñado para obtener un método estándar de colocación de los electrodos en estudios relativos a EEG. Dicho sistema se basa en la relación entre la localización de un electrodo y la superficie subyacente de la

corteza cerebral. El 10 y 20 se refieren al hecho de que las distancias reales entre los electrodos adyacentes están a 10 % o 20 % de la distancia total de delante-detrás o derecha-izquierda del cráneo. Para establecer las coordenadas cartesianas en dicho sistema, se utilizaron las coordenadas de todos los lugares de cada electrodo en una superficie de la cabeza *real* (con dimensiones reales) simulada en un programa de cómputo, sobre la base de las distancias a lo largo de la superficie (triangular) de la cabeza. La superficie de la cabeza utilizada fue construida a partir de la resonancia magnética canónica que se incluye en el paquete *SPM2*, y las ubicaciones se expresan en coordenadas *MNI* (standard Montreal Neurological Institute) (Oostenveld and Praamstra (2001)).

En la figura 4.1 se presenta la gráfica de las coordenadas de los 19 electrodos bajo el sistema 10-20.

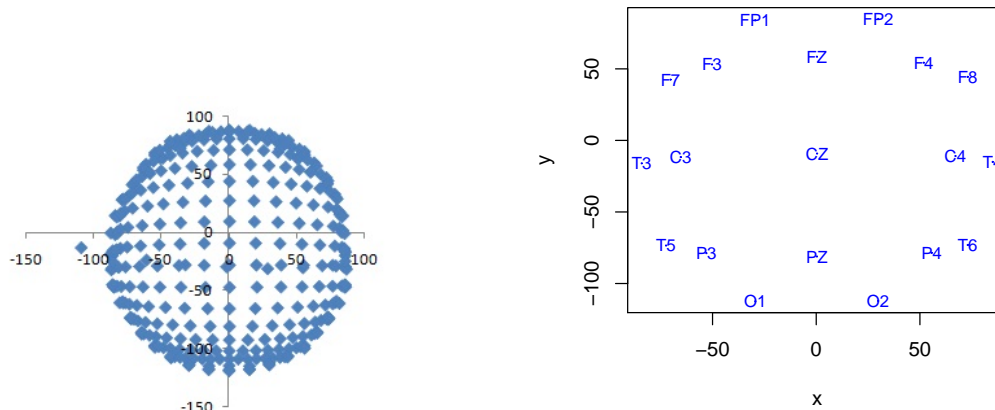


Figura 4.1.: Coordenadas en el sistema 10-20.

4.2. Análisis de varianza funcional sin estructura de dependencia espacial.

En esta sección se fija una ubicación espacial (electrodo), razón por la cual, los métodos aplicados asumen que las curvas son independientes (en éste caso los pacientes son considerados como independientes debido a que se trabaja en una ubicación espacial determinada y por lo tanto no hay estructura de correlación espacial). Se comparan las palabras con carga emocional positiva y negativa. Se realiza un análisis de varianza funcional clásico. Los parámetros funcionales son estimados utilizando la propuesta de Ramsay and Silverman (2005) (sección

2.3.1). Las pruebas ANOVA funcionales aplicadas corresponden a la F tiempo a tiempo propuesta por Ramsay and Silverman (2005) (sección 2.3.1) y a la versión sin correlación del test de Cuevas and Fraiman (2004) (sección 2.3.2), dado que éste ANOVA solo presenta un factor con dos niveles (carga emocional de la palabra).

4.2.1. Análisis de varianza funcional clásico para los hombres

A continuación se presentan los resultados del análisis de varianza funcional, teniendo en cuenta la información de los 30 hombres en la ubicación del electrodo $C3$ (de coordenadas $-65.358, -11.63$), dado que éste es uno de los electros más estudiados en la literatura sobre datos cerebrales. Este electrodo está ubicado en el lóbulo central del hemisferio izquierdo del cerebro. Sin embargo, se destaca que éste análisis puede ser realizado en cualquier otro electrodo.

En éste caso el modelo ajustado para comparar las palabras de carga emocional positiva y negativa en un electrodo determinado (previamente fijado) fue:

$$y_{ig}(t) = \mu(t) + \alpha_g(t) + \varepsilon_{ig}(t), \quad (4.1)$$

dónde $\alpha_g(t)$, $i = 1, 2, \dots, 30$. son cada uno de los 30 pacientes hombres del experimento y $g = 1, 2$. representan el tipo de carga emocional de la palabra mostrada al paciente.

Se desea probar la hipótesis nula:

$$H_0 : \alpha_1(t) = \alpha_2(t), \quad (4.2)$$

dónde $\alpha_1(t)$ es el efecto de palabras con carga emocional positiva y $\alpha_2(t)$ es el efecto de palabras con carga emocional negativa.

Suavizado de las curvas del electroencefalograma mediante B-splines.

Como se mencionó en los aspectos metodológicos, el primer paso para realizar los análisis subsecuentes, es establecer la base de funciones necesaria para suavizar los datos. En este caso se utilizó una base de B-splines debido a que los datos no presentaron patrones regulares de ciclo.

En la figura 4.2 se presenta el gráfico de validación cruzada para las observaciones de los pacientes hombres en la ubicación del electrodo $C3$. Se destaca que el número óptimo de funciones de la base fue 50.

Una vez suavizados cada uno de los 30 conjuntos de datos, éstos son incluidos como variable respuesta en un análisis de varianza funcional clásico (presentado en la sección 2.3.1). Los

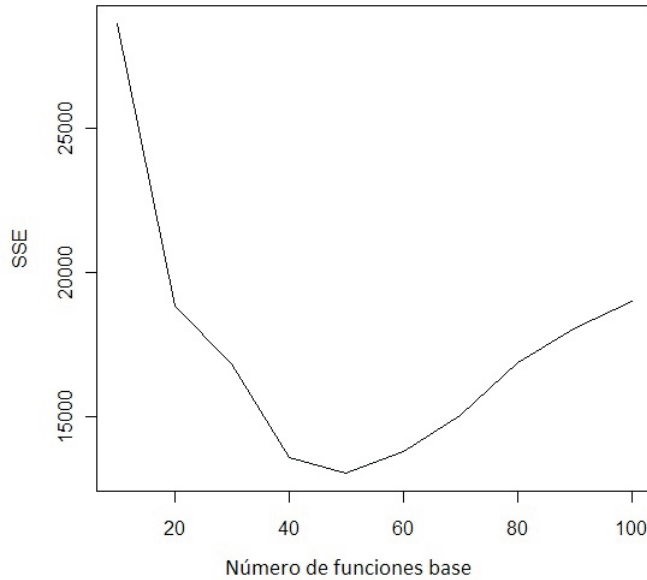


Figura 4.2.: Suma de cuadrados del error (SSE) obtenida por validación cruzada en función del número de funciones de la base.

tratamientos están determinados por el tipo de carga que presenta la palabra mostrada a cada paciente (ecuación (2.10)).

En el panel izquierdo de la figura 4.3 se muestran las curvas suavizadas mediante la base de 50 funciones B-splines para las respuestas de los 30 hombres en la ubicación del electrodo *C3* según el estímulo presentado: estímulo positivo (rojo) y negativo (azul). Este número (50) fue establecido mediante criterios de validación cruzada no paramétrica. El factor de penalización utilizado fue cero (0). En el panel derecho de la figura 4.3 se muestra la curva media (rojo) y de desviación estándar (verde) de dichas observaciones.

La curva media (mostrada en rojo en el panel derecho de la figura 4.3) sugiere que antes de la estimulación (milisegundo 200), la energía que emite el cerebro en la ubicación del electrodo *C3* es igual a la emitida en la ubicación del electrodo base (electrodo ubicado cerca a la oreja). Sin embargo, aproximadamente 100 milisegundos después de la estimulación (milisegundo 300), la energía emitida en el electrodo *C3* es mayor a la emitida en el electrodo base y éste efecto va diluyéndose al pasar el tiempo. Así al final de la curva (800 milisegundos después de la estimulación) se observa que la energía emitida en el electrodo *C3* es menor a la emitida en el electrodo base.

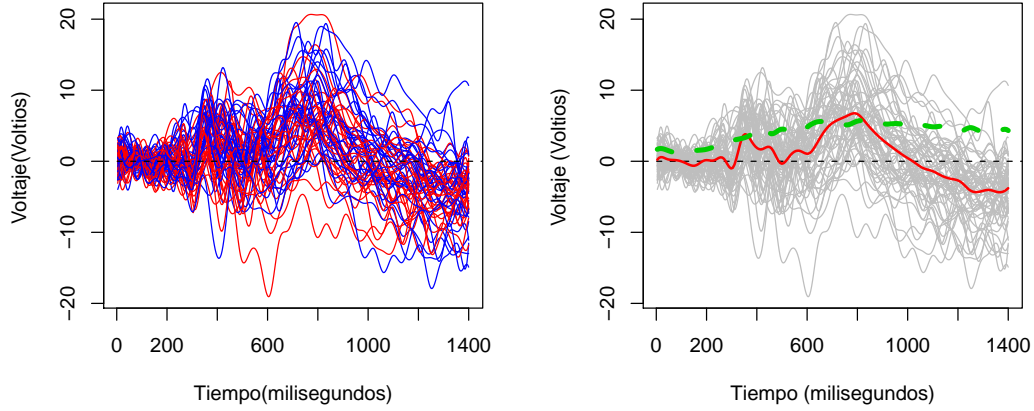


Figura 4.3.: Curvas suavizadas de los 30 hombres para el estímulo positivo (rojo) y para el estímulo negativo (azul) en la ubicación del electrodo C3 (Izquierda). Curva media y de desviación estándar (Derecha).

Estimación de los parámetros funcionales

Los parámetros funcionales del modelo de ANOVA de la ecuación (4.1) se estimaron utilizando la propuesta de Ramsay and Silverman (2005) (sección 2.3.1) mediante mínimos cuadrados ordinarios. Estos indican, que la energía que emite el cerebro de los hombres en la ubicación del electrodo C3 al ser estimulado con una palabra de carga positiva, es menor que la emitida en la ubicación de referencia (electrodo base). Situación contraria a la presentada cuando es estimulado con una palabra de carga negativa (figura 4.4). En cuanto a las medias funcionales estimadas (figura 4.5) se evidencia que el voltaje máximo (pico máximo) del estímulo positivo aparece antes del voltaje máximo del estímulo negativo. Esto indica que los hombres reaccionan más rápidamente a estímulos positivos que negativos. Sin embargo éste pico no se da en tiempos cercanos a los 300 milisegundos después de haber sido estimulado. Este aparece en tiempos cercanos a los 500 milisegundos después, lo que muestra que la tarea que ejecuta el paciente es compleja y por tal motivo la latencia del potencial es mayor a 300 milisegundos.

Utilizando la prueba F clásica para probar la hipótesis de la ecuación (4.2) (izquierda de la figura 4.6) en cada uno de los 350 tiempos (propuesta de Ramsay and Silverman (2005)), se puede concluir que con un 5 % de significancia, existen diferencias significativas en la cantidad de energía liberada por el cerebro de los hombres al ser objeto de estímulos positivos y negativos a partir de los 600 milisegundos en que aparece el estímulo y hasta 800 milisegun-

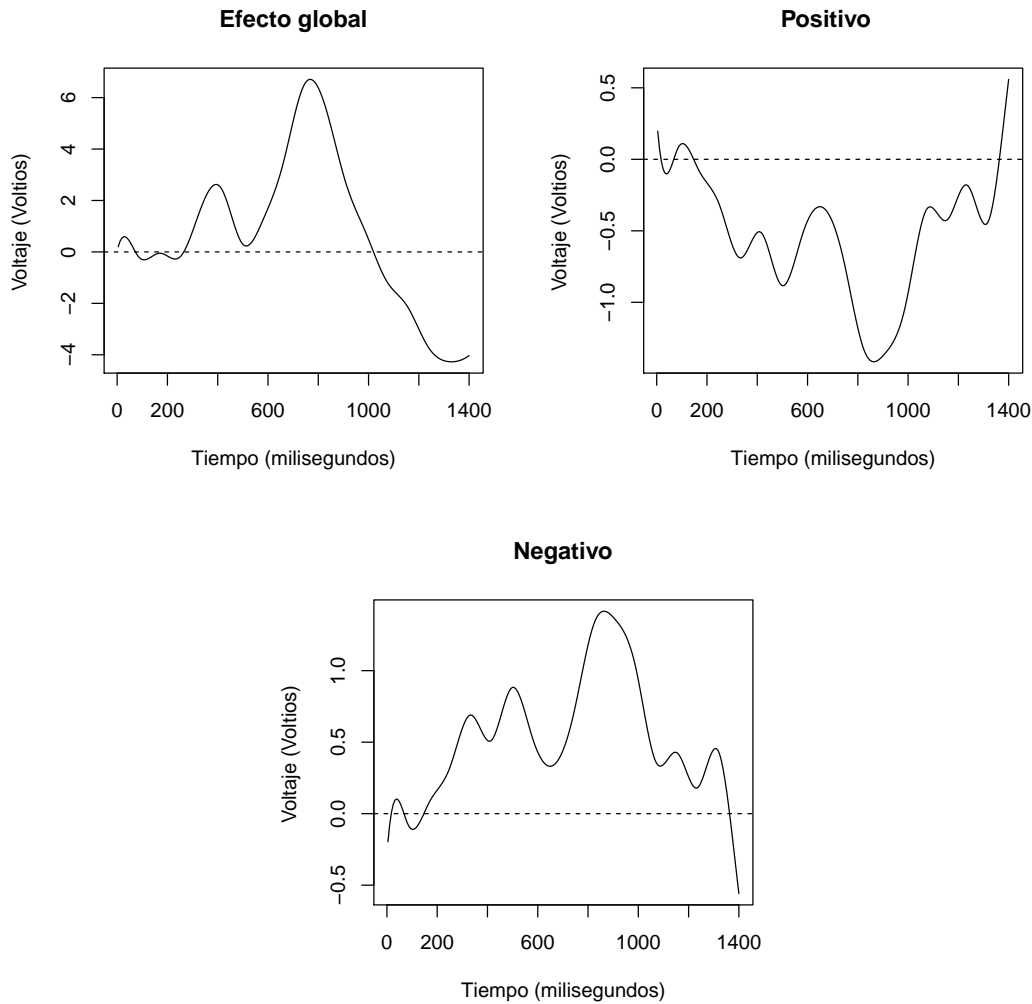


Figura 4.4.: Parámetros funcionales estimados para el análisis de varianza funcional de los hombres en la ubicación del electrodo C3.

dos. Sin embargo aplicando la prueba propuesta por Cuevas and Fraiman (2004) (sección 2.3.2), con un 5 % de significancia se concluye que no existen diferencias significativas entre los dos estímulos (derecha de la figura 4.6). Como se mencionó anteriormente la prueba F clásica tiempo a tiempo no tiene en cuenta la estructura de correlación temporal y por tal motivo es más conservador aplicar la propuesta de Cuevas and Fraiman (2004).

4.2.2. Análisis de varianza para las mujeres

Análogamente se presentan los resultados del análisis de varianza funcional clásico, teniendo en cuenta la información de las 18 mujeres en la ubicación del electrodo C3 (de coordenadas

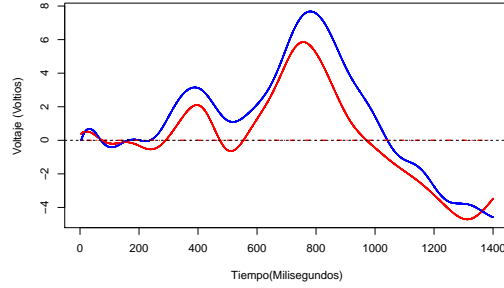


Figura 4.5.: Estimación de la curva media respuesta para los hombres con el estímulo positivo (rojo) y estímulo negativo (azul).

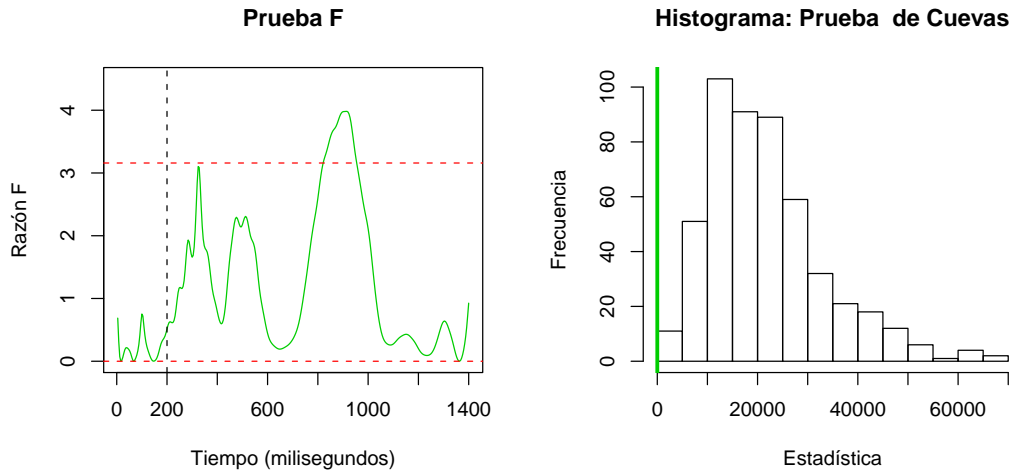


Figura 4.6.: Prueba F para la diferencia entre los estímulos en cada tiempo (Izquierda) y prueba funcional de Cuevas and Fraiman (2004) (Derecha).

$(-65.358, -11.63)$). Se destaca que éste análisis puede ser extrapolado a cualquier otro electrodo muestreado.

En este caso el modelo ajustado para comparar las palabras de carga emocional positiva y negativa en un electrodo determinado (previamente fijado) fue:

$$y_{ig}(t) = \mu(t) + \alpha_g(t) + \varepsilon_{ig}(t), \quad (4.3)$$

dónde $\alpha_g(t)$, $i = 1, 2, \dots, 18$. son cada una de las 18 pacientes mujeres del experimento y $g = 1, 2$. representan el tipo de carga emocional de la palabra mostrada al paciente.

Se desea probar la hipótesis nula:

$$H_0 : \alpha_1(t) = \alpha_2(t), \quad (4.4)$$

dónde $\alpha_1(t)$ es el efecto de palabras con carga emocional positiva y $\alpha_2(t)$ es el efecto de palabras con carga emocional negativa.

Suavizado de las curvas del electroencefalograma mediante B-splines

En el caso de las mujeres el número óptimo de funciones base estimado por validación cruzada fue 40. La curva media de la respuesta bio-eléctrica del cerebro de las mujeres en el electrodo *C3* presenta un comportamiento similar al de la curva de la función media de ésta respuesta en los hombres (ver figura 4.7).

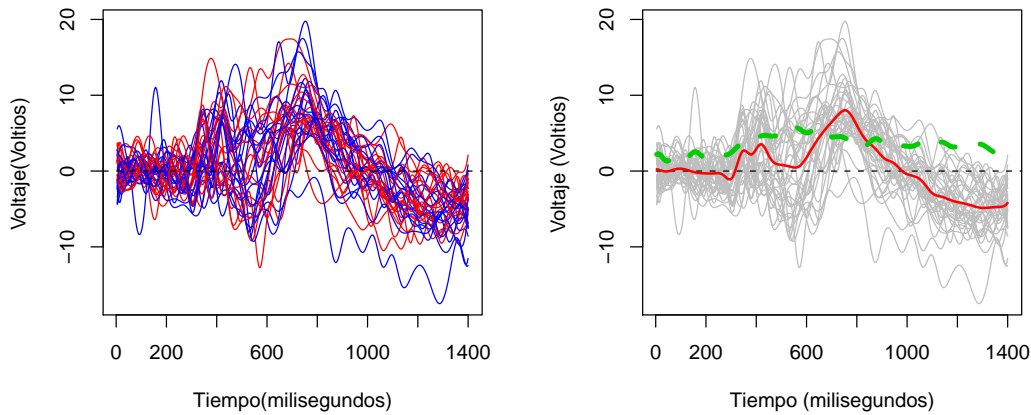


Figura 4.7.: Curvas suavizadas de las 18 mujeres para el estímulo positivo (rojo) y para el estímulo negativo (azul) en la ubicación del electrodo *C3*.

Estimación de los parámetros funcionales

Los parámetros funcionales del modelo ANOVA de la ecuación (4.3) se estimaron utilizando la propuesta de Ramsay and Silverman (2005) (sección 2.3.1). Estos muestran (figura 4.8), que la energía que emite el cerebro de las mujeres en la ubicación del electrodo *C3* al ser estimulado con una palabra de carga positiva y al ser estimulado con una palabra de carga negativa presenta un comportamiento opuesto en forma de ondas con una fluctuación alrededor de cero. La energía emitida por el cerebro en el electrodo *C3* es similar a la energía emitida en el electrodo base. Al aplicar la prueba F tiempo a tiempo (sección 2.3.1) para probar la hipótesis planteada en la ecuación (4.4), se concluye que con un 5% de significancia, no existen diferencias significativas en la cantidad de energía liberada por el cerebro de las mujeres al ser objeto de estímulos positivos y negativos (izquierda de la figura

4.10). Análogamente utilizando la prueba de Cuevas and Fraiman (2004) (sección 2.3.2) y con 5% de significancia la conclusión es similar (derecha de la figura 4.10). Se destaca que la respuesta máxima del cerebro de las mujeres a estímulos con carga emocional positiva se presenta antes de la respuesta máxima a estímulos de carga emocional negativa (figura 4.9), es decir, las mujeres reaccionan más rápidamente a estímulos positivos que negativos.

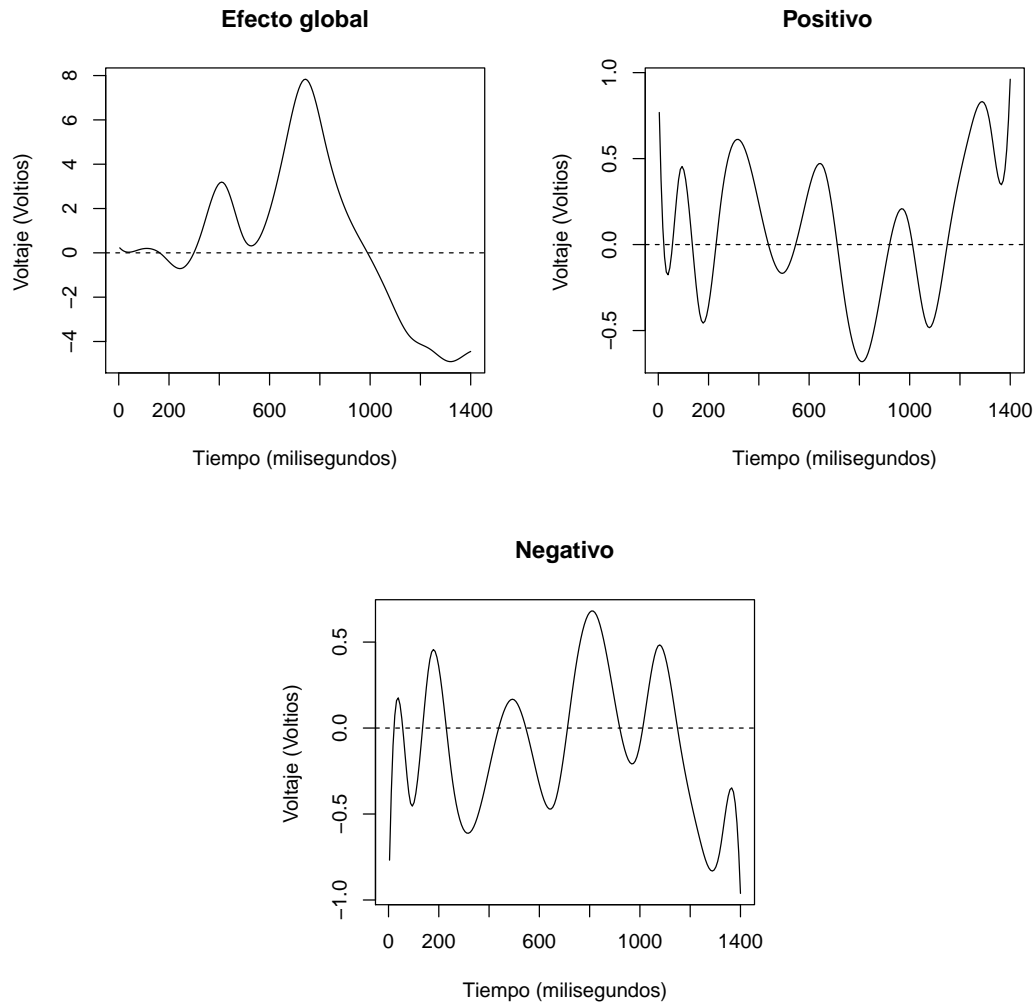


Figura 4.8.: Parámetros funcionales estimados para el análisis de varianza funcional de las mujeres en el electrodo *C3*.

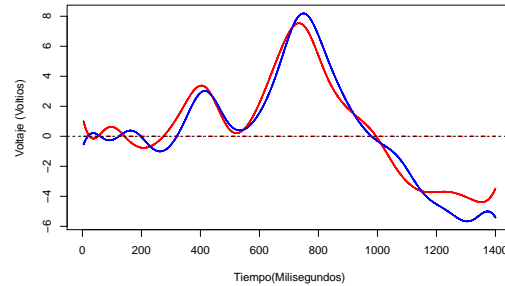


Figura 4.9.: Estimación de la curva media respuesta para las mujeres con el estímulo positivo (rojo) y estímulo negativo (azul).

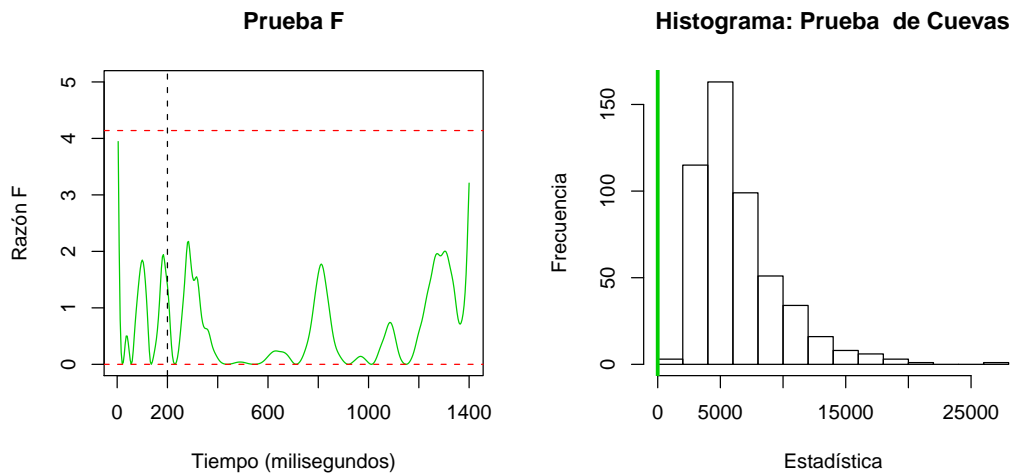


Figura 4.10.: Prueba F para la diferencia entre los estímulos (Izquierda) y prueba funcional de Cuevas and Fraiman (2004) (Derecha)

4.3. Análisis de varianza a dos vías de clasificación

Como se vió anteriormente, la energía liberada por el cerebro al ser estimulado con palabras de distinta carga emocional es diferente si el paciente es hombre o mujer. En el caso de la localización del electrodo *C3* (aunque éste análisis puede ser extrapolado a cualquier otro electrodo), los hombres presentan diferencias en algunos tiempos específicos, mientras que las mujeres no. En ésta sección se realiza un análisis de varianza con dos factores fijos que corresponden al tipo de estímulo al cual es sometido el paciente y al género del mismo. Se aplica un ANOVAF clásico a dos vías de clasificación. Las inferencias se realizan a través de la prueba de Shen and Faraway (2004) debido a que ésta incorpora varios factores de clasificación.

En términos formales, el modelo queda descrito como:

$$y_{igk}(t) = \mu(t) + \alpha_g(t) + \beta_k(t) + \varepsilon_{igk}(t), \quad (4.5)$$

dónde $\alpha_g(t)$, $g = 1, 2$. representan el tipo de carga emocional de la palabra mostrada al paciente, $\beta_k(t)$, $k = 1, 2$. representa el sexo del paciente y $i = 1, 2, \dots, n_g$ son cada uno de los pacientes del experimento estimulado con palabras de carga emocional g .

Se desea probar las hipótesis nulas:

$$H_0 : \alpha_1(t) = \alpha_2(t), \quad (4.6)$$

dónde $\alpha_1(t)$ es el efecto de palabras con carga emocional positiva y $\alpha_2(t)$ es el efecto de palabras con carga emocional negativa.

$$H_0 : \beta_1(t) = \beta_2(t), \quad (4.7)$$

dónde $\beta_1(t)$ corresponde al efecto del género masculino y $\beta_2(t)$ al femenino.

En éste caso el proceso de suavizamiento se realizó con 40 funciones de una base B-splines, dado que éste fue el número óptimo a través de validación cruzada. La curva media (figura 4.11) de todas las observaciones realizadas en la localización del electrodo *C3* establece que el voltaje máximo de liberación de energía se presenta alrededor del milisegundo 400, es decir 200 milisegundos después de la aparición del estímulo.

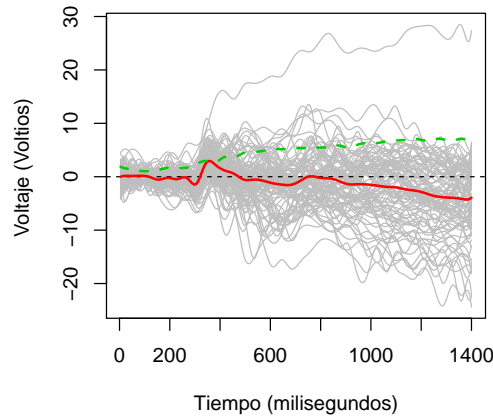


Figura 4.11.: Estimación de la curva respuesta media para todos los pacientes en la ubicación del electrodo *C3*.

Las funciones medias estimadas del modelo dado en la ecuación (4.5) a través de los procedimientos clásicos enunciados en la sección 2.3.1 mediante mínimos cuadrados ordinarios,

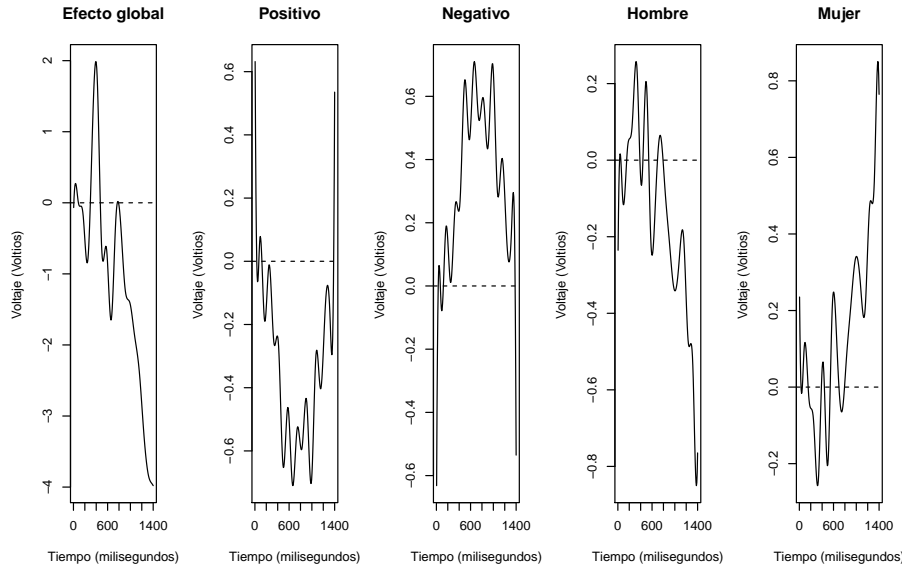


Figura 4.12.: Parámetros funcionales estimados para el ANOVA funcional de los pacientes en el electrodo *C3*

Tabla 4.1.: Prueba F funcional de Shen and Faraway (2004) para el ANOVAF propuesto en la ecuación (4.5).

Factor	Valor de la F calculada	Valor de la F teórica
Estímulo (ecuación (4.6))	0.183	1.1091
Género (ecuación (4.7))	0.194	1.1091

sugieren que los hombres estimulados con palabras de carga positiva emiten menos energía y las mujeres estimuladas con palabras de carga negativa emiten mayor energía (figura 4.11). Al aplicar el test F funcional propuesto por Shen and Faraway (2004) para probar las hipótesis de las ecuaciones (4.6) y (4.7), se concluye con un nivel de significancia del 5% que no existen diferencias significativas en la respuesta bio-eléctrica del cerebro de mujeres y hombres cuando son estimulados con palabras de distinta carga emocional (positivos versus negativos). En la tabla 4.1 se muestran los resultados de dicha prueba.

4.4. Análisis de varianza con datos correlacionados espacialmente

En ésta sección se tiene en cuenta la información de un paciente en las 19 ubicaciones espaciales (electrodos), es decir, se presenta una estructura de dependencia espacial entre las curvas. Se aplica la metodología propuesta en la sección 3.2 para estimar la estructura de correlación

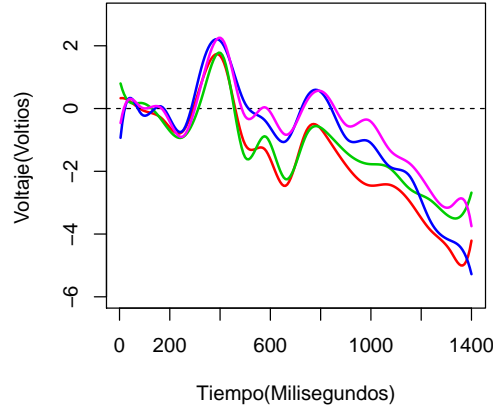


Figura 4.13.: Estimación de la curva respuesta para los hombres con estímulo positivo (rojo), hombres con estímulo negativo (azul), mujeres con estímulo positivo (verde) y mujeres con estímulo negativo (morado) en el electrodo C3.

espacial de las curvas de un mismo paciente en las 19 ubicaciones espaciales (electrodos). Se involucra dicha estructura en la estimación de los parámetros funcionales aplicando la metodología propuesta en 3.3. Finalmente ésta estructura es incluida en el cálculo de los residuales con los cuales se realizan las respectivas inferencias aplicando la metodología propuesta en la sección 3.4 y se realiza la aplicación de la propuesta en 3.5. En éste caso se presentan los resultados para un hombre y una mujer en particular. La selección se realizó de forma aleatoria. Se destaca que el análisis se puede establecer para cualquier paciente

El modelo lineal aplicado fue:

$$y_{ij}(t) = \mu(t) + \alpha_j(t) + \varepsilon_{ij}(t), \quad (4.8)$$

dónde $\alpha_i(t)$, $i = 1, 2, \dots, 19$. son cada una de los 19 electrodos y $j = 1, 2$. representan el tipo de carga emocional de la palabra mostrada al paciente.

Se desea probar la hipótesis nula:

$$H_0 : \alpha_1(t) = \alpha_2(t) \quad (4.9)$$

dónde $\alpha_1(t)$ es el efecto de palabras con carga emocional positiva y $\alpha_2(t)$ es el efecto de palabras con carga emocional negativa.

4.4.1. Paciente hombre

A continuación se presentan los resultados de las metodologías propuestas en el capítulo 3, considerando la información del paciente hombre identificado con el número 5. Este paciente

fue seleccionado de forma aleatoria de los 30 pacientes hombres en el estudio.

Después de realizar los procedimientos de validación cruzada (sección 2.2.2) el número óptimo de funciones de una base B-spline fue 50. En el panel izquierdo de la figura 4.14 se presentan las curvas suavizadas y en el derecho, las funciones medias para cada estímulo.

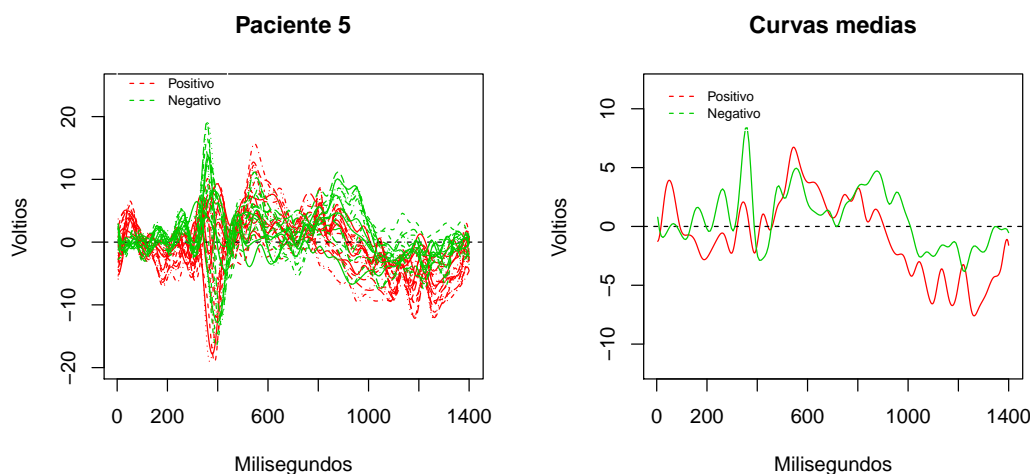


Figura 4.14.: Suavizamiento de funciones a través de B-splines para el paciente hombre identificado con el número 5 (Izquierda). Curva media de la respuesta bioeléctrica del cerebro al ser estimulado con palabras positivas (rojo) y negativas (verde) (Derecha).

Con las curvas suavizadas y con la matriz de distancias generada a través de las coordenadas de los 19 electrodos establecidas en el sistema de referencias 10-20, se realizó la estimación de las traza variogramas simples (con las curvas de cada uno de los tratamientos, ecuación (2.38)) y de la traza variograma cruzada (ecuación (3.2)). En la gráfica 4.15 se muestran las respectivas estimaciones. Los puntos en rojo representan a la traza variograma “*bin*”, es decir, la estimación realizada con los promedios de las diferencias para cada distancia h , mientras que los negros la estimación de la traza variograma “*cloud*”, es decir, solo las diferencias para los puntos distanciados por h .

Ajustando uno de los modelos conocidos a las estimaciones de las traza variogramas (secciones 2.1 y 3.2), el modelo que mejor ajusta éstas observaciones es un modelo exponencial de rango 130. En la tabla 4.2 se presentan los parámetros del modelo obtenidos y en la figura 4.16 los respectivos ajustes. Una vez estimada dicha estructura se calculó la matriz de varianzas y covarianzas integrada a partir de la ecuación (3.3).

En cuánto a la estimación de los parámetros funcionales del modelo lineal de la ecuación (4.8), se involucró la estructura de correlación espacial anteriormente estimada. Se usó la

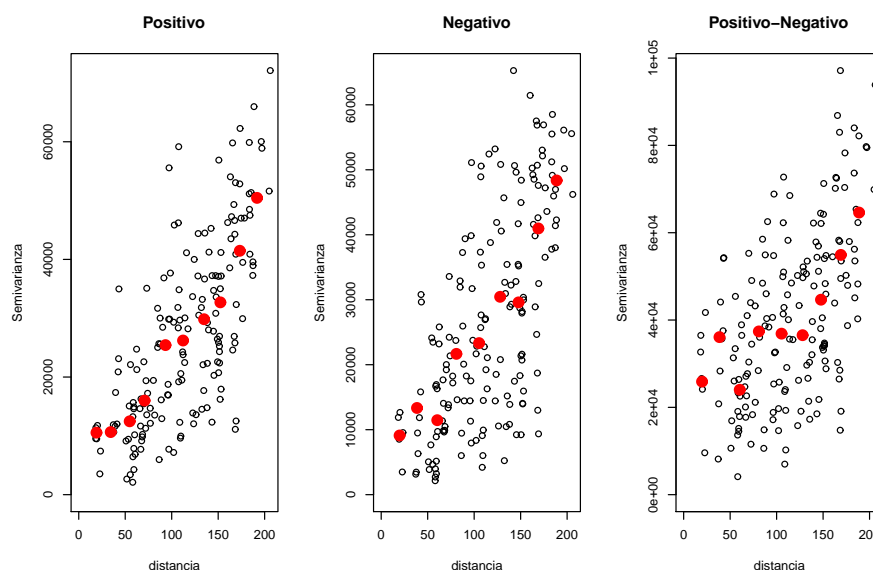


Figura 4.15.: Estimaciones de las traza variogramas y de la traza variograma cruzada para el paciente hombre identificado con el número 5. Los puntos rojos presentan la traza variograma bin y los negros la traza variograma cloud.

Tabla 4.2.: Parámetros estimados del modelo de las traza variogramas mediante MLC para el paciente hombre identificado con el número 5.

Variable	Silla	Rango
Estímulo positivo	64911.53	130
Estímulo negativo	64014.34	130
Estímulo positivo-negativo	63823.14	130

propuesta realizada en la sección 3.3 mediante mínimos cuadrados generalizados (ecuaciones (3.5) a (3.7)). Se destaca que en éste paciente, la energía liberada por el cerebro al ser objeto de palabras con carga emocional positiva es menor a la energía liberada cuando es estimulado con palabras de carga negativa. La latencia de ambos estímulos se produce 200 milisegundos después de haber sido estimulado (ver figura 4.17). Sin embargo, y aunque se mantiene dicha desigualdad, al estimar los parámetros sin involucrar esta correlación (con un análisis de varianza funcional clásico), la latencia del estímulo negativo se produce después de 100 milisegundos y la del estímulo positivo después de 300 milisegundos (ver figura 4.18). Al aplicar las pruebas ANOVA para probar la hipótesis descrita en la ecuación (4.9) con las modificaciones propuestas en el capítulo 3 (secciones 3.4 y 3.5), con un 5 % de significancia se concluye que existen diferencias significativas en la respuesta bio-eléctrica de éste paciente al ser estimulado con palabras de carga positiva y negativa (tabla 4.3).

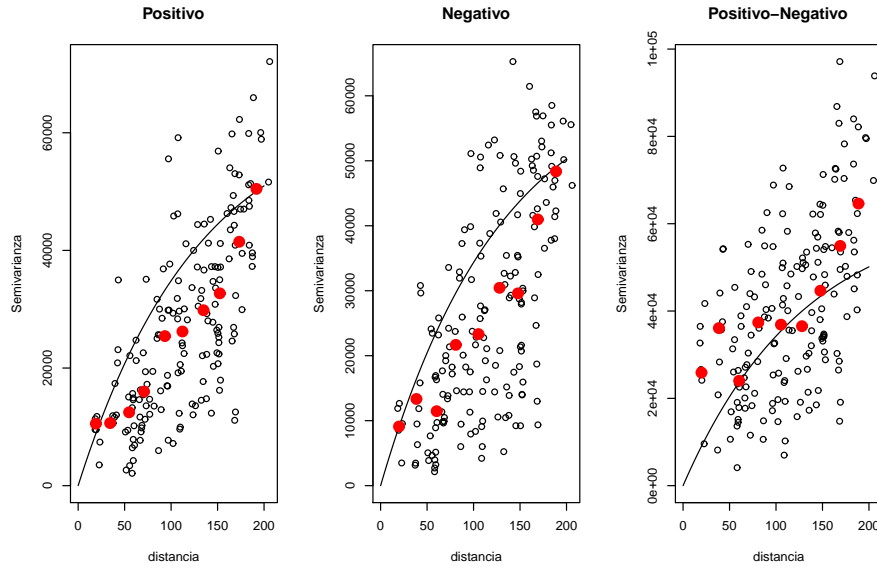


Figura 4.16.: Ajuste del modelo exponencial a las traza variogramas del paciente hombre identificado con el número 5.

Tabla 4.3.: Pruebas ANOVA para el paciente hombre identificado con el número 5.

Prueba	Con correlación	Sin correlación
Shen y Faraway	2.126*	4.311*
Cuestas-Albertos y Febrero-Bande	2.457*	18.459*

4.4.2. Paciente mujer

Para el caso de las mujeres, la mujer seleccionada aleatoriamente y para la cual se presentan los resultados, corresponde a la identificada con el número 2. Se destaca que éstos resultados pueden establecerse para cualquier mujer en el estudio. El suavizamiento fue realizado mediante una base de B-splines con 50 funciones (éste número de funciones se escogió nuevamente por validación cruzada). En el panel izquierdo de la figura 4.19 se presentan las curvas suavizadas y a la derecha, las funciones medias para cada estímulo (rojo para palabras de carga positiva y verde para palabras de carga negativa).

Con las curvas suavizadas y con la matriz de distancias generada a través de las coordenadas de los 19 electrodos, se realizó la estimación de las traza variogramas (ecuación (2.38)) para las curvas de cada uno de los dos estímulos y de la respectiva traza variograma cruzada (ecuación (3.2)). En la gráfica 4.20 se muestran las respectivas estimaciones. Los puntos en rojo representan la traza variograma “*bin*”, es decir, la estimación realizada con los promedios de las diferencias para cada distancia h mientras que los negros la “*cloud*”, es decir, solo las diferencias para los puntos distanciados por h .

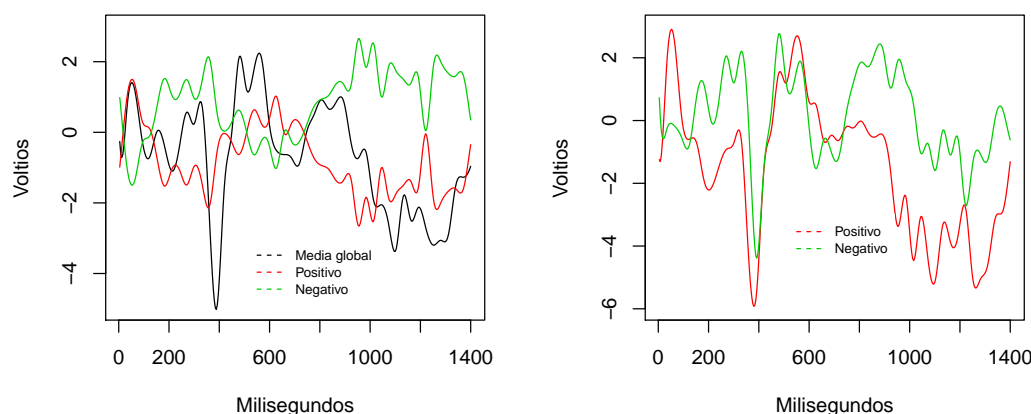


Figura 4.17.: Estimación de los parámetros funcionales involucrando la estructura de correlación espacial para el paciente hombre identificado en el número 5 (Izquierda). Estimación de las medias funcionales (Derecha).

Tabla 4.4.: Parámetros estimados del modelo de las traza variogramas mediante MLC para la paciente mujer identificada con el número 2.

Variable	Silla	Rango
Estímulo positivo	77451.78	100
Estímulo negativo	64564.95	100
Estímulo positivo-negativo	70015.27	100

Ajustando uno de los modelos conocidos a las estimaciones de las traza variogramas (secciones 2.1 y 3.2), el modelo que mejor ajusta estas observaciones es un modelo exponencial de rango 100. En la tabla 4.4 se presentan los parámetros del modelo obtenidos y en la figura 4.21 los respectivos ajustes. Una vez estimada dicha estructura se calculó la matriz de varianzas y covarianzas integrada a partir de la ecuación (3.2).

En cuanto a la estimación de los parámetros funcionales del modelo lineal de la ecuación (4.8), se involucró la estructura de correlación espacial anteriormente estimada. Se usó la propuesta realizada en la sección 3.3 mediante mínimos cuadrados generalizados (ecuaciones (3.5) a (3.7)). Se destaca que la energía liberada por el cerebro al ser objeto de palabras con carga emocional positiva es mayor a la energía liberada cuando es estimulado con palabras de carga negativa. La latencia de ambos estímulos se produce 300 milisegundos después de ser estimulado (ver figura 4.22). Al estimar los parámetros funcionales sin involucrar dicha correlación, las conclusiones son similares. Sin embargo (ver figura 4.23) en este caso, al

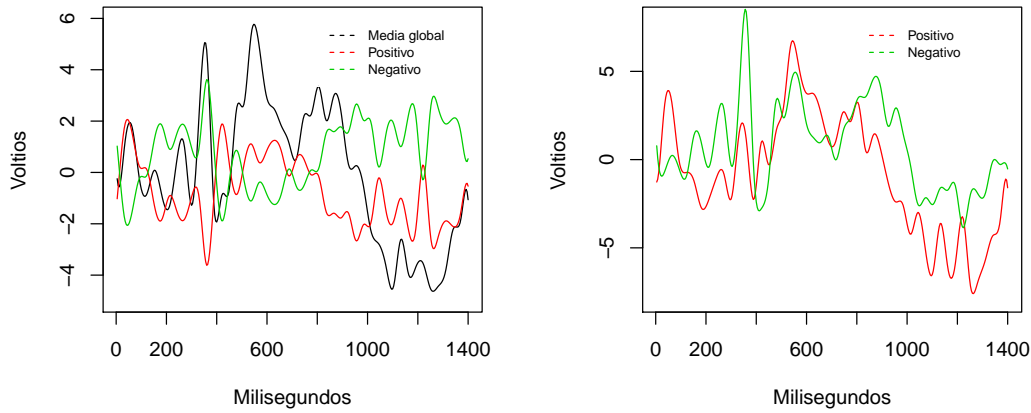


Figura 4.18.: Estimación de los parámetros funcionales sin involucrar la estructura de correlación espacial para el paciente hombre identificado en el número 5 (Izquierda). Estimación de las medias funcionales (Derecha).

Tabla 4.5.: Pruebas ANOVA para la paciente mujer identificada con el número 2.

Prueba	Con correlación	Sin correlación
Shen y Faraway	0.667	1.704*
Cuestas-Albertos y Febrero-Bande	0.422	4.619*

aplicar las pruebas ANOVA para probar la hipótesis descrita en la ecuación (4.9) (secciones 3.4 y 3.5), con un 5 % de significancia se concluye que existen diferencias significativas en la respuesta bio-eléctrica de ésta paciente al ser estimulada con palabras de carga positiva y negativa, conclusión contraria a la establecida si ésta estructura es considerada (tabla 4.5).

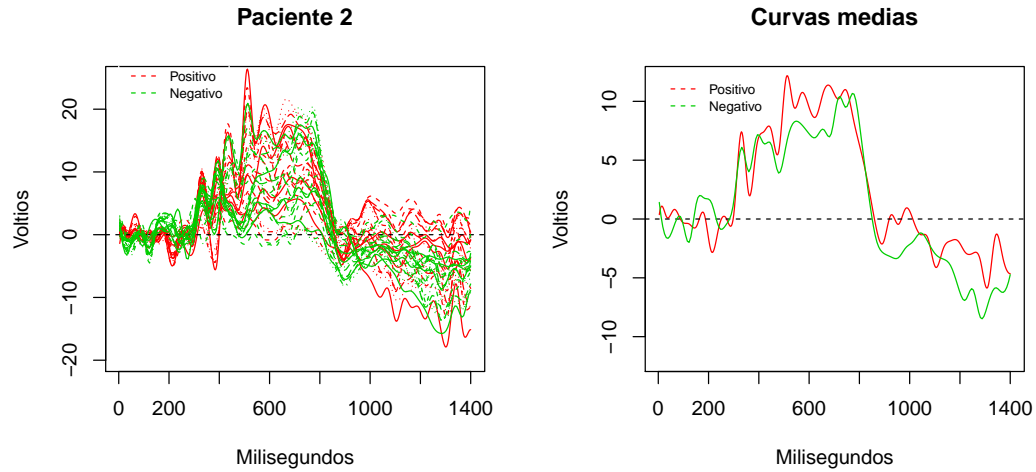


Figura 4.19.: Suavizamiento de funciones a través de B-splines para la paciente mujer identificada con el número 2 (Izquierda). Curva media de la respuesta bio-eléctrica del cerebro al ser estimulado con palabras positivas (rojo) y negativas (verde) (Derecha).

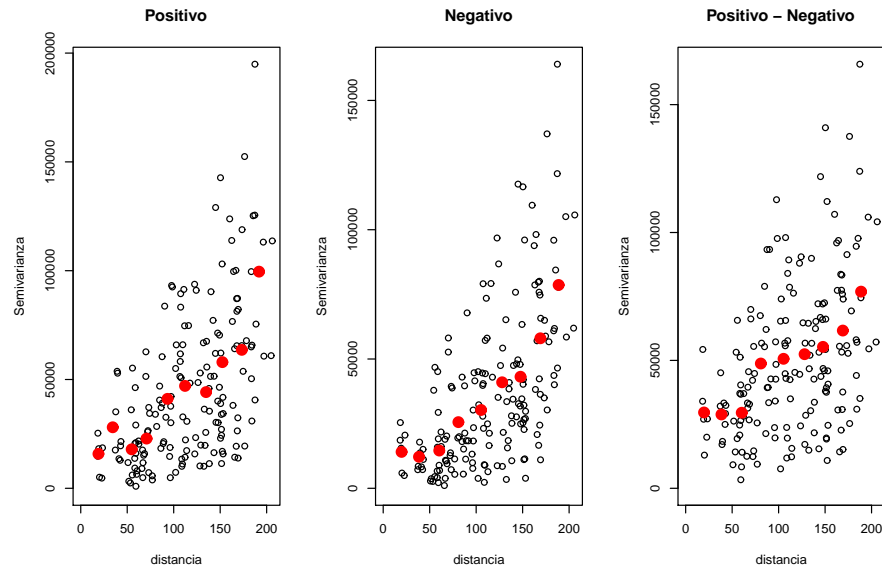


Figura 4.20.: Estimaciones de las traza variogramas y de la traza variograma cruzada para la paciente mujer identificada con el número 2. Los puntos rojos presentan la traza variograma bin y los negros la traza variograma cloud.

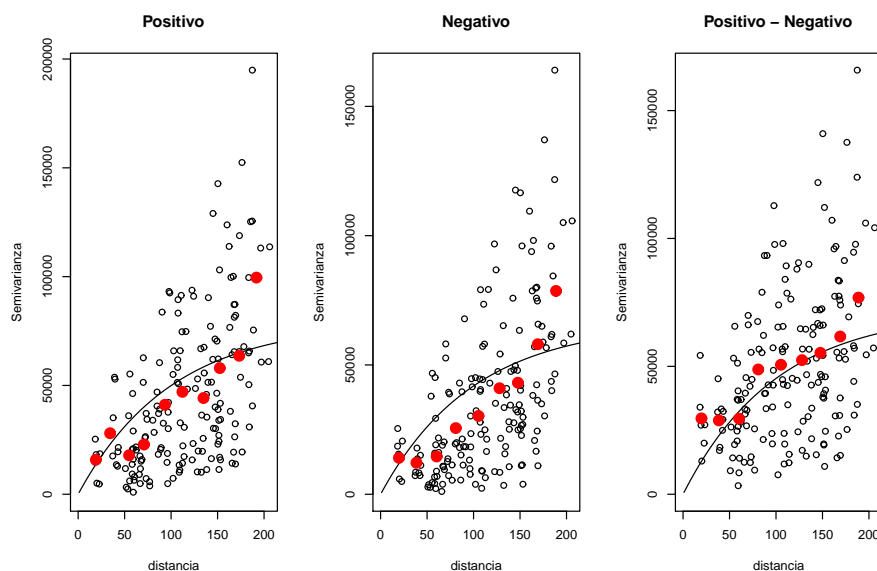


Figura 4.21.: Ajuste del modelo exponencial a las traza variogramas de la paciente mujer identificada con el número 2.

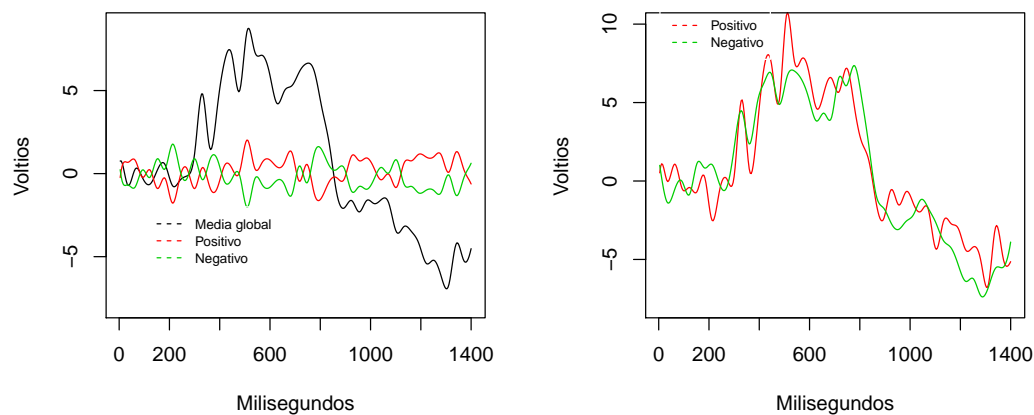


Figura 4.22.: Estimación de los parámetros funcionales involucrando la estructura de correlación espacial para la paciente mujer identificada en el número 2 (Izquierda). Estimación de las medias funcionales (Derecha).

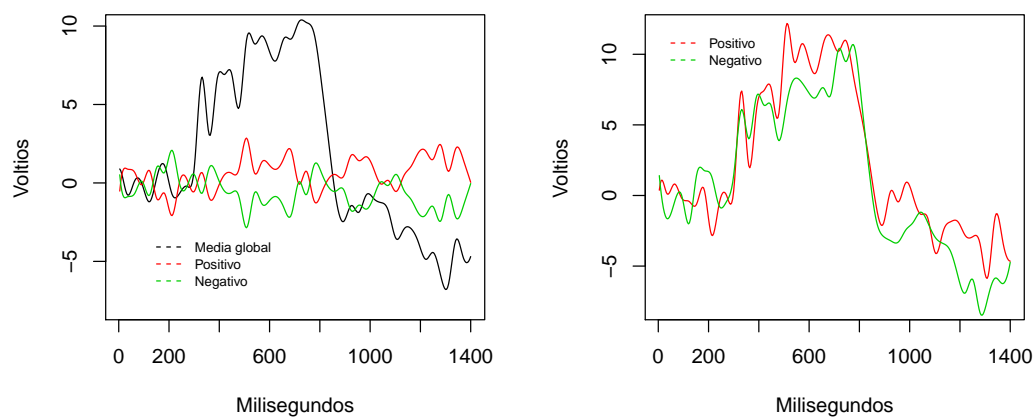


Figura 4.23.: Estimación de los parámetros funcionales sin involucrar la estructura de correlación espacial para la paciente mujer identificada en el número 2 (Izquierda). Estimación de las medias funcionales (Derecha).

5. Conclusiones

- Se propone una metodología de análisis de varianza funcional involucrando una estructura de correlación espacial. Se extiende la metodología de estimación de los parámetros del ANOVA funcionales (ANOVAF) propuesta por Ramsay and Silverman (2005) a través de mínimos cuadrados generalizados.
- Se hace una generalización de la traza variograma al caso de dos variables aleatorias funcionales y se propone su uso en la estimación de la estructura de covarianza espacial en modelos ANOVA funcionales (ANOVAF).
- Se realiza la extensión de la prueba ANOVA funcional propuesta por Shen and Faraway (2004), al caso en el que las curvas residuales presentan estructura de correlación espacial.
- Se realiza la extensión de la prueba ANOVA funcional de Cuesta-Albertos and Febrero-Bande (2007), al caso en el que las curvas residuales presentan estructura de correlación espacial.
- Las metodologías propuestas son *plug-in*, debido a que se anidan varios procesos de estimación. Por un lado, la estimación de la estructura de correlación espacial entre curvas y por el otro la inclusión de ésta en el proceso de estimación de los parámetros y pruebas de un ANOVA funcional.
- Es importante destacar que al realizar el ajuste de la traza del variograma simple y cruzado, el modelo que mejor ajusta no siempre es el mismo, por lo cual la etapa de diagnóstico de éste se convierte en parte fundamental del proceso.
- Los algoritmos establecidos para el desarrollo completo de ésta tesis fueron modificaciones de rutinas programadas en R. Se encuentran a disposición del usuario.

6. Trabajo futuro

- Realizar estudios de simulación con el fin de determinar las propiedades estadísticas distribucionales de las pruebas extendidas.
- Establecer mecanismos que permitan realizar una estimación conjunta de todos los parámetros involucrados, desde la etapa inicial de suavizamiento (número de funciones base y parámetros de penalidad), hasta la etapa final de estimación de parámetros funcionales.
- Realizar aplicaciones a otras áreas con el fin determinar posibles comportamientos en situaciones extremas.
- Finalmente la validación de supuestos sobre las curvas residuales (distribución, independencias, homocedasticidad, linealidad) de un modelo como el propuesto es un campo de poco estudio. Este podría ser un tema interesante a explorar.

A. Funciones en R para realizar ANOVAF con errores independientes

En este capítulo se encuentran los programas utilizados para realizar las aplicaciones de los modelos de las secciones 4.2 y 4.3. Se destaca que en estos casos, se considera la información de los pacientes en un electrodo determinado, razón por la cual, estas curvas se consideran independientes. En la primera parte se encuentra el programa para estimar los parámetros de un ANOVA funcional considerando la información de los pacientes de un género determinado (hombres o mujeres) en una ubicación espacial específica (electrodo). Se realizan las pruebas F tiempo a tiempo y la prueba propuesta por Cuevas and Fraiman (2004). En la segunda parte se considera la información de todos los pacientes (hombres y mujeres) en la misma ubicación. Se estiman los parámetros funcionales y se realiza la prueba de Shen and Faraway (2004) para el caso de un ANOVAF a dos vías (se compra la respuesta bio-eléctrica del cerebro en diferentes estímulos y géneros). Los códigos pueden obtenerse de la página [http : //www.docentes.unal.edu.co/rgiraldoh/docs/](http://www.docentes.unal.edu.co/rgiraldoh/docs/).

A.1. Código para ANOVAF de los pacientes de un género determinado en un electrodo determinado

```
#####
# Código para hacer Regresión Funcional con
# los datos de un electrodo determinado, con los
# pacientes de un genero determinado, los 2 estímulos
#(palabras positivas y negativas),para un total de
# 58 filas) y 350 mediciones en el tiempo.
#####
#####
# Librerías necesarias
#####

rm(list=ls())
library(zoo)
library(fda)
library(geofd)
library(fda.usc)
```

```
#####
# Lectura de datos
#####

datos<-read.table("BrainDT.txt",header = T, dec = ",")
levels(datos$esite)=c("C3","C4","CZ","F3","F4","F7","F8",
                     "FP1","FP2","FZ","O1","O2","P3","P4",
                     "PZ","T3","T4","T5","T6","VEOG")

#####
# Creación de la base, determinando el sexo de los
# pacientes y el electrodo que se quiere estudiar
#####

BASEPOSI<-(datos$esite=="C3")&(datos$gender=="fema")&
            (datos$wordtype=="1")

BASENEGA<-(datos$esite=="C3")&(datos$gender=="fema")&
            (datos$wordtype=="2")

BASEPOSI = as.matrix(datos[BASEPOSI,4:353])
BASENEGA = as.matrix(datos[BASENEGA,4:353])

BASETOTAL =(datos$esite=="C3")&(datos$gender=="fema")&
            ((datos$wordtype=="1")|(datos$wordtype=="2"))

BASETOTAL =as.matrix(datos[BASETOTAL,4:353])
BASETOTALT = t(BASETOTAL)
t = dim(BASETOTALT)[1]
n = dim(BASETOTALT)[2]

BASEPOSIT = t(BASEPOSI)
dim(BASEPOSIT)
BASENEGAT = t(BASENEGA)
dim(BASENEGAT)

#####
# Creación de una base B-splines de orden 4 y k = 40.
#####

time<- c( seq(4,1400,4))
rangeval = c(4,1400)
nbasis = 40
norder = 4
basis6 = create.bspline.basis(rangeval, nbasis, norder)
```

```

col <-matrix(2, nrow=1, ncol=n)

par(mfrow=c(1,2))
fdhposi = data2fd(BASEPOSIT, time, basis6)
plot(fdhposi,col=1, xlab="Tiempo", ylab="Voltaje", lty=1,
     pch = 56, main = "Estímulo positivo" )
fdhnega = data2fd(BASENEGAT, time, basis6)
plot(fdhnega, xlab="Tiempo", ylab="Voltaje", lty=1,
     pch = 56, col = 4, main = "Estímulo negativo" )
fdh = data2fd(BASETOTALT, time, basis6)
plot(fdh,col=rep(c(1,4), n), xlab="Tiempo(milisegundos)",
     ylab="Voltaje(Voltios)", lty=1,pch = 56)

#####
#Determinación y gráfica de la función media y de varianza.
#####

meanfdh <- mean.fd(fdh)
varfdh <-var.fd(fdh)
stdvfdh <- stddev.fd(fdh)
par(mfrow=c(1,2))
plot(meanfdh, main="Media", xlab="Tiempo", ylab="Voltaje" )
plot(stdvfdh, main="Desviación estándar", xlab="Tiempo", ylab="Voltaje")
plot(stdvfdh^2, main="Varianza", xlab="Tiempo", ylab="Voltaje")

windows()
plot(fdh,col=8, xlab="Tiempo (milisegundos)", ylab="Voltaje (Voltios)", lty=1)
lines(meanfdh,col=2,lwd=2)
lines(stdvfdh, ylim=c(0,10), main="Desviación estándar",
     xlab="Tiempo (milisegundos)", ylab="Voltaje (Voltios)",lty=2,
     col=3, lwd=2)

#####
# Modelo lineal con variable respuesta curva de
# la energia liberada por el cerebro y variable
# explicativa el tipo de estímulo al cual el paciente
# es sometido.
#####

matrizh <- eval.basis(time, basis6)
y2cMap <- solve(crossprod(matrizh)) %*% t(matrizh)

#####
#Nombre de los estímulos

```

```
#####

estimulos <- c("Efecto global", "Positivo ", "Negativo")

#####
#Indices para los estimulos
#####

posiindex <- seq(1,n,2)
negindex <- seq(2,n,2)
#####
#Matriz diseño
#####
p = 3
zmat <- matrix(0,n,p)
zmat[,1] <- 1
zmat[posiindex,2] <- 1
zmat[negindex,3] <- 1

#####
#Generación de la fila, para garantizar la ortogonalidad
#y estimabilidad de los parámetros.
#Fila (0,1,1,1)
#####

z59 <- matrix(1,1,p)
z59[1] <- 0
zmat <- rbind(zmat, z59)

#####
#Estimación de la fila generada a través de las
#funciones base.
#Fila (0,1,1,1).
#####

coef <- fdh$coefs
coef59 <- cbind(coef,matrix(0,nbasis,1))
fdh$coefs <- coef59

#####
# Estimación de los parametros funcionales, a través
# de 20 funciones base y de orden 4
#####

xhdlst <- vector("list",p)
for (j in 1:p) xhdlst[[j]] <- zmat[,j]
```

```

nbetabasis <- 20
betabasis = create.bspline.basis(rangeval, nbetabasis, norder)

betaafd <- fd(matrix(0,nbetabasis,1), betabasis)
estimate <- T
lambda <- 0
betaafdPar <- fdPar(betaafd)

betalist <- vector("list",p)
for (j in 1:p) betalist[[j]] <- betaafdPar

fRegressList <- fRegress(fdh, xfdlist, betalist)

#####
# Gráfica de parámetros funcionales (betas) estimados
#####

betaestlist <- fRegressList$betaestlist
par(mfrow=c(3,1))
for (j in 1:p)
{
  betaestParfdj <- betaestlist[[j]]
  plot(betaestParfdj$fd, xlab="Tiempo (milisegundos)",
       ylab="Voltaje (Voltios)",
       main=estimulos[j])
}

#####
# Gráfica de funciones predichas
#####

yfdpar <- fRegressList$yfdpar

yhatfdobj <- fRegressList$yhatfdobj
plot(yhatfdobj)

#####
# Matriz del modelo reducido
#####

MatrizMod = matrix(rep(1),n,1)

#####

```

```
# Estimación del parametro funcional del modelo reducido
# a través de 20 funciones base y de orden 4
#####
fdhred = data2fd(BASETOTALT, time, basis6)
q <- 1
xfdlistRed <- vector("list",q)
for (j in 1:q) xfdlistRed[[j]] <- MatrizMod[,j]

nbetabasis <- 20
betabasis = create.bspline.basis(rangeval, nbetabasis, norder)

betafd <- fd(matrix(0,nbetabasis,1), betabasis)
estimate <- T
lambda <- 0
betafdPar <- fdPar(betafd)

betalistRed <- vector("list",q)
for (j in 1:q) betalistRed[[j]] <- betafdPar

fRegressListRed <- fRegress(fdhred, xfdlistRed, betalistRed)

#####
# Gráfica de parámetros funcionales (betas) estimados
# para el modelo reducido
#####

betaestlistRed <- fRegressListRed$betaestlist
for (j in 1:q)
{
  betaestParfdj <- betaestlistRed[[j]]
  plot(betaestParfdj$fd, xlab="Tiempo (milisegundos)",
        ylab="Voltaje (Voltios)",
        main=estimulos[j])
}

#####
# Gráfica de funciones predichas para el modelo reducido
#####

yfdparRed <- fRegressListRed$yfdpar

yhatfdobjRed<- fRegressListRed$yhatfdobj
plot(yhatfdobjRed)
```

```
#####
# Cálculo de la matriz de residuales y obtención de
# la matriz de varianza y covarianza
#####

yhatmat <- eval.fd(time,yhatfdobj$fd)      # Valores estimados
matplot(yhatmat[,1:n], col=2, lty=1, lwd=2)
ymat <- eval.fd(time, fdh)                # Valores observados
matplot(time,ymat[,1:n], type="l")
alfa <- betaestlistRed[[1]]
alpha <- eval.fd(time,alfa$fd)            # Constantes (Media) estimada
plot(alfa$fd, xlab="Tiempo", ylab="Voltaje")
Residuales <- ymat[,1:n] - yhatmat[,1:n]   # Los residuales son iguales
                                           # a observados - estimados

matplot(Residuales, type = "l")

#####
# Cálculo de las sumas de cuadrados
#####

SCT<-matrix(0, nrow=t, ncol=n) #Suma de cudrados Total
for(i in 1:n)
{
  SCT [,i] <-(ymat[,i]-alpha)^2
}
SCT<-apply(SCT,1,sum)
plot(time,SCT, type="l", col=3, main="Suma de cuadrados total")

SCE <- apply(Residuales^2, 1 , sum) #Suma de cudrados del error

plot(time,SCE,type="l", col=3, main="Suma de cuadrados del error")

SCR<-matrix(0, nrow=t, ncol=n) #Suma de cudrados de la regresión

for(i in 1:n)
{
  SCR [,i] <-(yhatmat[,i]-alpha)^2
}
SCR<-apply(SCR,1,sum)
plot(time, SCR, col=3, main="Suma de cuadrados de la regresión",
      ylim = c(0,350))

SCR1 = SCT-SCE
plot(time, SCR1, col=3, type="l")
```

```
#####
# Cálculo de los cuadrados medios y de la prueba F
# tiempo a tiempo
#####

dferror =n - p
dfRegresion = p-q
CME<-SCE/dferror
plot(CME, col =2)
CMR<-SCR/dfRegresion
CMR1<-SCR1/dfRegresion

plot(CME, col = 2)
plot(CMR, col = 2)

F<-CMR/CME
plot(time,F, main="Prueba F", col = 3, xlab="Tiempo (milisegundos)",
ylab="Razón F", type="l", ylim=c(0,4.5))
Fteo<-qf(0.95,dfRegresion,dferror)
abline(h=Fteo, lty=2, col="red")
abline(h=0, lty=2, col="red")
abline(v=200, col=1, lty=2)

#####
# Cálculo de la estadística de Cuevas#####
#####
estimulo0 = betaestlist[[1]]
estimulo1 = betaestlist[[2]]
estimulo2 = betaestlist[[3]]

media = eval.fd(time,betaestlist[[1]]$fd)
est1 = eval.fd(time,betaestlist[[2]]$fd)
est2 = eval.fd(time,betaestlist[[3]]$fd)

estimulo1 = media + est1
estimulo2 = media + est2

vectorCom = cbind(estimulo1,estimulo2)
vectorCom =as.matrix(vectorCom)

ndif<-dim(vectorCom)[2]
time<- c( seq(4,1400,4))
nbasis = 50
norder = 4
smooth.type = "bsplines"
lambda = 0
```



```
vectorMedF= .create.fd.object(vectorCom, smooth.type, time, nbasis, lambda)
vectorMedF$datafd$coef
```

```
CuevT= (n/2)*12.norm(ndif,vectorMedF$datafd,vectorMedF$M)[1,2]
```

```
#####
# Gráfica de contorno de la varianza de los residuales
#####
```

```
SigmaE <- var(t(Residuales))
contour(SigmaE, xlab="Milisegundos", ylab="Milisegundos", cex=1.2)
```

```
#####
# Gráfica de dsviación estandar de los residuales
#####
cexval = 1.2
par(mfrow=c(1,1), mar=c(5,5,4,2)+cexval+2, pty="m")
stddevE <- sqrt(diag(SigmaE))
plot(time, stddevE, type="l", cex=1.2,
      xlab="Milisegundos", ylab="Desviación estandar (Voltios)")
```

```
#####
# Estimación de los coeficientes de regresión con
# intervalos de confianza.
#####
```

```
stderrList <- fRegress.stderr(fRegressList, y2cMap, SigmaE)
betastderrlist <- stderrList$betastderrlist
```

```
#####
# Gráfica de los coeficientes de regresión con
# intervalos de confianza.
#####
```

```
par(mfrow=c(2,3))
for (j in 1:p) {
  betastderrj <- eval.fd(time, betastderrlist[[j]])
  plot(time, betastderrj, cex=1.2,
        type="l",lty=1, xlab="Milisegundos", ylab="Reg. Coeff.",
        main=estimulos[j])
}
```

```
par(mfrow=c(2,3))
for (j in 1:p) {
  betafdParj <- betaestlist[[j]]
```

```

betafdj      <- betafdParj$fd
betaj        <- eval.fd(time, betafdj)
betastderrj  <- eval.fd(time, betastderrlist[[j]])
matplot(time, cbind(betaj, betaj+2*betastderrj, betaj-2*betastderrj),
         type="l",lty=c(1,4,4), xlab="Milisegundos", ylab="Reg. Coeff.",
         main=estimulos[j])
}

#####
#Programa de bostrap para realizar
#inferencias a través de la estadística de Cuevas
#####

#####
#Librerías necesarias
#####
library(fda)
library(geoR)
library(geofd)
library(gstat)

#####
#Escogencia de la base de datos
#####
BASETOTALA =(datos$esite=="C3")&(datos$gender=="fema")&((datos$wordtype=="1")|
               (datos$wordtype=="2"))
BASETOTAL  =as.matrix(datos[BASETOTALA,4:353])
BASETOTAL2 =as.matrix(datos[BASETOTALA,3])
BASETOTALT = t(BASETOTAL)
dim(BASETOTALT)
dim(BASETOTALT)

#####
# Procedimiento bootstrap
#####
rep = 100
Cuev = rep(0,rep)

for(k in 1:rep)
{
t = dim(BASETOTALT)[1]
n = dim(BASETOTALT)[2]
tam = n
x = rep(0,tam)
es = seq(1:n)
m = sample(es,tam,replace = TRUE, prob = NULL)
mat = matrix(rep(0,tam*t),t,tam)
trat = rep(0,tam)

```

```

for(i in 1:tam)
{
mat[,i] = BASETOTALT[,m[i]]
trat[i] = BASETOTAL2[m[i]]
}
matfin = t(mat)
MatrizMod = matrix(rep(1),tam,3)
MatrizMod[,2] = (trat-1)
MatrizMod[,3] = (1-MatrizMod[,2])

time<- c( seq(4,1400,4))
rangeval = c(4,1400)
nbasis = 40
norder = 4
basis6 = create.bspline.basis(rangeval, nbasis, norder)

fdmat = data2fd(t(matfin), time, basis6)

#####
#Estimación de los parámetros
#####
#####
#Nombre de los estímulos
#####

estimulos <- c("Efecto global", "Positivo ", "Negativo")

#####
#Generación de la fila, para garantizar la ortogonalidad
#y estimabilidad de los parámetros.
#Fila (0,1,1)
#####

ztam <- matrix(1,1,3)
ztam[1] <- 0
zmat <- rbind(MatrizMod, ztam)

#####
#Estimación de la fila generada a través de las
#funciones base.

```

```
#Fila (0,1,1).
#####

coef  <- fdmat$coefs
coeftam <- cbind(coef,matrix(0,nbasis,1))
fdmat$coefs <- coeftam

#####
# Estimación de los parametros funcionales, a través
# de 20 funciones base y de orden 4
#####

p <- 3
xfdlist <- vector("list",p)
for (j in 1:p) xfdlist[[j]] <- zmat[,j]

nbetabasis <- 20
betabasis = create.bspline.basis(rangeval, nbetabasis, norder)

betafd  <- fd(matrix(0,nbetabasis,1), betabasis)
estimate <- T
lambda  <- 0
betafdPar <- fdPar(betafd)

betalist <- vector("list",p)
for (j in 1:p) betalist[[j]] <- betafdPar

fRegressList <- fRegress(fdmat, xfdlist, betalist)

#####
# Gráfica de parámetros funcionales (betas) estimados
#####

betaestlist <- fRegressList$betaestlist
par(mfrow=c(3,1))
for (j in 1:p)
{
  betaestParfdj <- betaestlist[[j]]
  plot(betaestParfdj$fd, xlab="Tiempo (milisegundos)", ylab="Voltaje (Voltios)",
       main=estimulos[j])
}

#####
# Cálculo de la estadística de Cuevas#####
```

```
#####
estimulo0 = betaestlist[[1]]
estimulo1 = betaestlist[[2]]
estimulo2 = betaestlist[[3]]

media = eval.fd(time,betaestlist[[1]]$fd)
est1 = eval.fd(time,betaestlist[[2]]$fd)
est2 = eval.fd(time,betaestlist[[3]]$fd)

estimulo1 = media + est1
estimulo2 = media + est2

vectorCom = cbind(estimulo1,estimulo2)
vectorCom =as.matrix(vectorCom)

ndif<-dim(vectorCom)[2]
time<- c( seq(4,1400,4))
nbasis = 50
norder = 4
smooth.type = "bsplines"
lambda = 0
vectorMedF= .create.fd.object(vectorCom, smooth.type, time, nbasis, lambda)
vectorMedF$datafd$coef

Cuev[k] = tam*l2.norm(ndif,vectorMedF$datafd,vectorMedF$M)[1,2]
}

hist(Cuev, main = "Histograma: Prueba de Cuevas", ylab = "Frecuencia",
      xlab = "Estadística")

abline(a = 0, b = CuevT, col =3, lwd = 3)
#####
```

A.2. Código para ANOVA en un electrodo determinado (a dos vias)

```
#####
# Código para hacer Regresión Funcional con
# los datos de un lectrodo determinado, con los
# pacientes de ambos genero determinado, los 2 estímulos
#(palabras positivas y negativas),para un total de
# 58 filas) y 350 mediciones en el tiempo.
#####

rm(list=ls())
library(zoo)
```

```

library(fda)

datos<-read.table("BrainDT.txt",header = T, dec = ",")

levels(datos$esite)=c("C3","C4","CZ","F3","F4","F7","F8",
                     "FP1","FP2","FZ","O1","O2","P3","P4",
                     "PZ","T3","T4","T5","T6","VEOG")

#####
# Creación de la base, determinando el sexo de los
# pacientes y el electrodo que se quiere estudiar
#####

BASEPOSI<-(datos$esite=="C3")&(datos$wordtype=="1")&
  (datos$subject !="1")

BASENEGA<-(datos$esite=="C3")&(datos$wordtype=="2")&
  (datos$subject !="1")
BASEPOSI = as.matrix(datos[BASEPOSI,4:353])
BASENEGA = as.matrix(datos[BASENEGA,4:353])

BASETOTAL =(datos$esite=="F8")&((datos$wordtype=="1")|
  (datos$wordtype=="2"))&(datos$subject !="1")

BASETOTAL =as.matrix(datos[BASETOTAL,4:353])
BASETOTALT = t(BASETOTAL)
t = dim(BASETOTALT)[1]
n = dim(BASETOTALT)[2]

BASEPOSIT = t(BASEPOSI)
dim(BASEPOSIT)
BASENEGAT = t(BASENEGA)
dim(BASENEGAT)

#####
# Creación de una base B spline de orden 4 y k = 40.
#####

time<- c( seq(4,1400,4))
rangeval = c(4,1400)
nbasis = 40
norder = 4
basis6 = create.bspline.basis(rangeval, nbasis, norder)

col <-matrix(2, nrow=1, ncol=n)

```

```

par(mfrow=c(1,2))
fdhposi = data2fd(BASEPOSIT, time, basis6)
plot(fdhposi,col=1, xlab="Tiempo", ylab="Voltaje", lty=1,
     pch = 56, main = "Estímulo positivo" )
fdhnega = data2fd(BASENEGAT, time, basis6)
plot(fdnega, xlab="Tiempo", ylab="Voltaje", lty=1,
     pch = 56, col = 4, main = "Estímulo negativo" )
fdh = data2fd(BASETOTALT, time, basis6)
plot(fdh,col=rep(c(2,4), n), xlab="Tiempo(milisegundos)",
     ylab="Voltaje(Voltios)", lty=1,pch = 56)

#####
#Determinación y grafica de la media y la varianza.
#####

meanfdh <- mean.fd(fdh)
varfdh <-var.fd(fdh)
stdvfdh <- stddev.fd(fdh)
par(mfrow=c(1,2))
plot(meanfdh, main="Media", xlab="Tiempo",ylab="Voltaje" )
plot(stdvfdh, main="Desviación estándar", xlab="Tiempo",ylab="Voltaje")
plot(stdvfdh^2, main="Varianza", xlab="Tiempo", ylab="Voltaje")

windows()
plot(fdh,col=8, xlab="Tiempo (milisegundos)", ylab="Voltaje (Voltios)",lty=1)
lines(meanfdh,col=2,lwd=2)
lines(stdvfdh, ylim=c(0,10), main="Desviación estándar",
     xlab="Tiempo (milisegundos)", ylab="Voltaje (Voltios)",
     lty=2, col=3, lwd=2)

#####
# Modelo lineal con variable respuesta la curva de
# la energia liberada por el cerebro y variable
# explicativa el tipo de estímulo al cual el paciente
# es sometido y el respectivo genero
#####

matrizh <- eval.basis(time, basis6)
y2cMap <- solve(crossprod(matrizh)) %*% t(matrizh)

#####
#Nombre de los estímulos
#####

```

```

estimulos <- c("Efecto global", "Positivo ", "Negativo",
              "Hombre", "Mujer")

#####
#Indices para los estimulos y los generos
#####

posiindex <- c(1,3,5,7,9,11,13,15,17,
              19, 21, 23,25,27,29,31,33,
              35, 37, 39, 41, 43, 45, 47, 49, 51,
              53, 55, 57)
negindex <- c(2,4,6,8,10,12,14,16,18,20,
              22,24,26,28,30,32,34,36, 38, 40,
              42, 44, 46, 48, 50, 52, 54, 56, 58)

hombres = seq(1,58,1)
mujeres = seq(59,92,1)
#####
#Matriz diseño del modelo completo
#####

zmat <- matrix(0,92,5)
zmat[,1] <- 1
zmat[posiindex,2] <- 1
zmat[negindex,3] <- 1
zmat[hombres,4] <- 1
zmat[mujeres,5] <- 1

#####
#Generación de la fila, para garantizar la ortogonalidad
#y estimabilidad de los parámetros referentes a estímulos.
#Fila (0,1,1)
#####

z93 <- matrix(1,1,5)
z93[1] <- 0
z93[4] <- 0
z93[5] <- 0
zmat <- rbind(zmat, z93)

#####
#Estimación de la fila generada a través de las
#funciones base.
#Fila (0,1,1).
#####

```



```

coef    <- fdh$coefs
coef93 <- cbind(coef,matrix(0,nbasis,1))
fdh$coefs <- coef93

#####
#Generación de las filas, para garantizar la ortogonalidad
#y estimabilidad de los parámetros referentes a genero.
#Fila (0,1,1)
#####

z94    <- matrix(1,1,5)
z94[1] <- 0
z94[2] <- 0
z94[3] <- 0
zmat   <- rbind(zmat, z94)

#####
#Estimación de la fila generada a través de las
#funciones base.
#Fila (0,1,1).
#####

coef    <- fdh$coefs
coef94 <- cbind(coef,matrix(0,nbasis,1))
fdh$coefs <- coef94

#####
# Estimación de los parametros funcionales, a través
# de 20 funciones base y de orden 4
#####

p <- 5
xfdlist <- vector("list",p)
for (j in 1:p) xfdlist[[j]] <- zmat[,j]

nbetabasis <- 20
betabasis = create.bspline.basis(rangeval, nbetabasis, norder)

betafd    <- fd(matrix(0,nbetabasis,1), betabasis)
estimate  <- T
lambda    <- 0
betafdPar <- fdPar(betafd)

betalist <- vector("list",p)

```

```

for (j in 1:p) betalist[[j]] <- betafdPar

fRegressList <- fRegress(fdh, xfdlist, betalist)

#####
# Gráfica de parámetros funcionales (betas) estimados
#####

betaestlist <- fRegressList$betaestlist
par(mfrow=c(1,5))
for (j in 1:p)
{
  betaestParfdj <- betaestlist[[j]]
  plot(betaestParfdj$fd, xlab="Tiempo (milisegundos)",
        ylab="Voltaje (Voltios)",
        main=estimulos[j])
}

BETA1 = betaestlist[[1]]
BETA2 = betaestlist[[2]]
BETA3 = betaestlist[[3]]
BETA4 = betaestlist[[4]]
BETA5 = betaestlist[[5]]

TRAT1 = BETA1$fd + BETA2$fd + BETA4$fd
TRAT2 = BETA1$fd + BETA2$fd + BETA5$fd
TRAT3 = BETA1$fd + BETA3$fd + BETA4$fd
TRAT4 = BETA1$fd + BETA3$fd + BETA5$fd

plot(TRAT1, col = 2, ylim = c(-6,3), xlab = "Tiempo(Milisegundos)",
      ylab = "Voltaje(Voltios)", lwd= 2)
lines(TRAT2, col = 3, lwd = 2)
lines(TRAT3, col = 4, lwd = 2)
lines(TRAT4, col = 6, lwd = 2)
windows()
#####
# Gráfica de funciones predichas
#####

yfdpar <- fRegressList$yfdpar

yhatfdobj <- fRegressList$yhatfdobj
plot(yhatfdobj$fd, xlab = "Tiempo(Milisegundos)",
      ylab = "Voltaje(Voltios)", lwd = 2)

#####
# Cálculo de la matriz de residuales y obtención de

```

```

# la matriz de varianza y covarianza
#####

yhatmat <- eval.fd(time,yhatfdobj$fd)      # Valores estimados
matplot(yhatmat[,1:92], col=2, lty=1, lwd=2)
ymat <- eval.fd(time, fdh)                # Valores observados
matplot(time,ymat[,1:92], type="l")
alfa <- betaestlist[[1]]
alpha <- eval.fd(time,alfa$fd)            # Constantes (Media) estimada
plot(alfa$fd, xlab="Tiempo", ylab="Voltaje")
Residuales <- ymat[,1:92] - yhatmat[,1:92] # Los residuales son iguales a
                                           # observados - estimados

matplot(Residuales, type = "l")

#####
# Gráfica de contorno de la varianza de los residuales del modelo completo
#####

SigmaE <- var(t(Residuales))
contour(SigmaE, xlab="Milisegundos", ylab="Milisegundos", cex=1.2)
dim(SigmaE)
#####
# Gráfica de dsviación estandar de los residuales del modelo completo
#####
cexval = 1.2
par(mfrow=c(1,1), mar=c(5,5,4,2)+cexval+2, pty="m")
stddevE <- sqrt(diag(SigmaE))
plot(time, stddevE, type="l", cex=1.2,
      xlab="Milisegundos", ylab="Desviación estandar (Voltios)")

#####
# Estimación del modelo reducido 1: Solo los estímulos
#####
#####
#Nombre de los estímulos
#####

estimulos <- c("Efecto global", "Positivo ", "Negativo")

#####
#Indices para los estímulos
#####

posiindex1 <- seq(1,92,2)

```

```

negindex1 <- seq(2,92,2)

#####
#Matriz diseño
#####
q1 = 3
fdh1 = data2fd(BASETOTALT, time, basis6)
zmat1 <- matrix(0,92,3)
zmat1[,1] <- 1
zmat1[posiindex1,2] <- 1
zmat1[negindex1,3] <- 1

#####
#Generación de la fila, para garantizar la ortogonalidad
#y estimabilidad de los parámetros.
#Fila (0,1,1,1)
#####

z931 <- matrix(1,1,3)
z931[1] <- 0
zmat1 <- rbind(zmat1, z931)

#####
#Estimación de la fila generada a través de las
#funciones base.
#Fila (0,1,1,1).
#####

coef <- fdh1$coefs
coef931 <- cbind(coef,matrix(0,nbasis,1))
fdh1$coefs <- coef931

#####
# Estimación de los parametros funcionales, a través
# de 20 funciones base y de orden 4
#####

xfdlist1 <- vector("list",q1)
for (j in 1:q1) xfdlist1[[j]] <- zmat1[,j]

nbetabasis <- 20
betabasis = create.bspline.basis(rangeval, nbetabasis, norder)

betafd <- fd(matrix(0,nbetabasis,1), betabasis)
estimate <- T
lambda <- 0

```

```

betafdPar <- fdPar(betafd)

betalist1 <- vector("list",q1)
for (j in 1:q1) betalist1[[j]] <- betafdPar

fRegressList1 <- fRegress(fdh1, xfdlist1, betalist1)

#####
# Gráfica de parámetros funcionales (betas) estimados
#####

betaestlist1 <- fRegressList1$betaestlist
par(mfrow=c(3,1))
for (j in 1:q1)
{
  betaestParfdj <- betaestlist1[[j]]
  plot(betaestParfdj$fd, xlab="Tiempo (milisegundos)",
        ylab="Voltaje (Voltios)",
        main=estimulos[j])
}

#####
# Gráfica de funciones predichas
#####

yfdpar1 <- fRegressList1$yfdpar

yhatfdobj1 <- fRegressList1$yhatfdobj
plot(yhatfdobj1)

#####
# Cálculo de la matriz de residuales y obtención de
# la matriz de varianza y covarianza del modelo reducido 1: Solo estimulos
#####

yhatmat1 <- eval.fd(time,yhatfdobj1$fd)      # Valores estimados
matplot(yhatmat1[,1:n], col=2, lty=1, lwd=2)
ymat1 <- eval.fd(time, fdh1)                 # Valores observados
matplot(time,ymat1[,1:n], type="l")
alfa <- betaestlist1[[1]]
alpha <- eval.fd(time,alfa$fd)               # Constantes (Media) estimada
plot(alfa$fd, xlab="Tiempo", ylab="Voltaje")
Residuales1 <- ymat1[,1:n] - yhatmat1[,1:n]  # Los residuales son iguales a
                                              # observados - estimados

matplot(Residuales1, type = "l")

```

```

SigmaRed1E  <- var(t(Residuales1))
contour(SigmaRed1E, xlab="Milisegundos", ylab="Milisegundos", cex=1.2)
dim(SigmaRed1E)
#####
# Gráfica de dsviación estandar de los residuales del modelo reducido
#####
cexval = 1.2
par(mfrow=c(1,1), mar=c(5,5,4,2)+cexval+2, pty="m")
stddevRed1E <- sqrt(diag(SigmaRed1E))
plot(time, stddevRed1E, type="l", cex=1.2,
      xlab="Milisegundos", ylab="Desviación estandar (Voltios)")

#####
# Prueba F de Fareway para comparar modelo completo menos modelo reducido
# Prueba sobre los géneros
#####

Num =( sum(diag(SigmaRed1E-SigmaE))/(p-q1))
Den = (sum(diag(SigmaE))/(n-p))
F1 = Num/Den
factor = (((sum(diag(SigmaE)))^2)/ (sum(diag(SigmaE^2))))
f1 = factor*(p-q1)
f2 = factor*(n-p)
Fteo1<-qf(0.95,f1,f2)

#####
# Estimación del modelo reducido 2: Solo los géneros
#####
#####
#Nombre de los estímulos
#####
q2 = 3
fdh2 = data2fd(BASETOTALT, time, basis6)
estimulos <- c("Efecto global", "Hombres", "Mujeres")

#####
#Indices para los estímulos
#####

Hombres1 <- seq(1,58,1)
Mujeres1 <- seq(59,92,1)

#####
#Matriz diseño
#####

```

```

zmat2 <- matrix(0,92,3)
zmat2[,1] <- 1
zmat2[Hombres1,2] <- 1
zmat2[Mujeres1,3] <- 1

#####
#Generación de la fila, para garantizar la ortogonalidad
#y estimabilidad de los parámetros.
#Fila (0,1,1,1)
#####

z932 <- matrix(1,1,3)
z932[1] <- 0
zmat2 <- rbind(zmat2, z932)

#####
#Estimación de la fila generada a través de las
#funciones base.
#Fila (0,1,1,1).
#####

coef <- fdh2$coefs
coef932 <- cbind(coef,matrix(0,nbasis,1))
fdh2$coefs <- coef932

#####
# Estimación de los parametros funcionales, a través
# de 20 funciones base y de orden 4
#####

xfdlist2 <- vector("list",q2)
for (j in 1:q2) xfdlist2[[j]] <- zmat2[,j]

nbetabasis <- 20
betabasis = create.bspline.basis(rangeval, nbetabasis, norder)

betafd <- fd(matrix(0,nbetabasis,1), betabasis)
estimate <- T
lambda <- 0
betafdPar <- fdPar(betafd)

betalist2 <- vector("list",q2)
for (j in 1:q2) betalist2[[j]] <- betafdPar

fRegressList2 <- fRegress(fdh2, xfdlist2, betalist2)

```

```
#####
# Gráfica de parámetros funcionales (betas) estimados
#####

betaestlist2 <- fRegressList2$betaestlist
par(mfrow=c(3,1))
for (j in 1:p)
{
  betaestParfdj <- betaestlist2[[j]]
  plot(betaestParfdj$fd, xlab="Tiempo (milisegundos)",
        ylab="Voltaje (Voltios)",
        main=estimulos[j])
}

#####
# Gráfica de funciones predichas
#####

yfdpar2 <- fRegressList2$yfdpar

yhatfdobj2 <- fRegressList2$yhatfdobj
plot(yhatfdobj2)

#####
# Cálculo de la matriz de residuales y obtención de
# la matriz de varianza y covarianza del modelo reducido 1: Solo géneros
#####

yhatmat2 <- eval.fd(time,yhatfdobj2$fd)      # Valores estimados
matplot(yhatmat2[,1:n], col=2, lty=1, lwd=2)
ymat2 <- eval.fd(time, fdh)                  # Valores observados
matplot(time,ymat2[,1:n], type="l")
alfa <- betaestlist2[[1]]
alpha <- eval.fd(time,alfa$fd)               # Constantes (Media) estimada
plot(alfa$fd, xlab="Tiempo", ylab="Voltaje")
Residuales2 <- ymat2[,1:n] - yhatmat2[,1:n]  # Los residuales son iguales a
                                              # observados - estimados

matplot(Residuales2, type = "l")

SigmaRed2E <- var(t(Residuales2))
contour(SigmaRed2E, xlab="Milisegundos", ylab="Milisegundos", cex=1.2)
dim(SigmaRed2E)
#####
# Gráfica de dsviación estandar de los residuales del modelo reducido
#####
```



```

cexval = 1.2
par(mfrow=c(1,1), mar=c(5,5,4,2)+cexval+2, pty="m")
stddevRed1E <- sqrt(diag(SigmaRed2E))
plot(time, stddevRed2E, type="l", cex=1.2,
      xlab="Milisegundos", ylab="Desviación estandar (Voltios)")

#####
# Prueba F de Fareway para comparar modelo completo menos modelo reducido
# Prueba sobre los estímulos
#####
Num =( sum(diag(SigmaRed2E-SigmaE))/(p-q2))
Den = (sum(diag(SigmaE))/(n-p))
F2 = Num/Den
factor = (((sum(diag(SigmaE)))^2)/ (sum(diag(SigmaE^2))))
f12 = factor*(p-q2)
f22 = factor*(n-p)
Fteo2<-qf(0.95,f12,f22)

```

B. Funciones en R para realizar ANOVAF con errores correlacionados

En este capítulo se presenta el programa utilizado para estimar los parámetros funcionales del ANOVA dado en la sección 4.4. En la primera parte se estima la estructura de covarianza y posteriormente esta es incluida en el proceso de estimación de los parámetros funcionales. Finalmente se encuentra la programación referente a la extensión de las pruebas de Shen and Faraway (2004) y Cuesta-Albertos and Febrero-Bande (2007). Los códigos pueden obtenerse de la página [http : //www.docentes.unal.edu.co/rgiraldoh/docs/](http://www.docentes.unal.edu.co/rgiraldoh/docs/).

B.1. Código para ANOVAF de un paciente determinado considerando las curvas en los 19 electrodos

```
#####  
# Código para determinar la traza variograma de un  
# paciente determinado con la información de los 19  
# electrodos del sistema 10-20  
#####  
#####  
# Librerías necesarias  
#####  
rm(list=ls())  
library(fda)  
library(geoR)  
library(geofd)  
library(gstat)  
library(fda.usc)  
#####  
# Paquetes modificados  
#####  
source("Norma_L2.R")  
source("TrazaModel.R")  
source("MatrizIzq.R")  
source("Principal.R")  
source("fit.Variogram1.R")  
source("lmc.R")
```

```

source("lmcNN.R")
source("lmcCN.R")
source("Trace.Vari1.R")
#####
# Lectura de los datos y coordenadas
#####
datos<-read.table("BrainDT.txt",header = T, dec = ",")
coordena = read.table("CoorFinales.txt",header = T, dec = ",")
plot(coordena[,2], coordena[,3], cex = 0.1, col = 4, xlab = "x", ylab = "y")
text(coordena[,2], coordena[,3], labels = coordena[,1], cex = 0.8, col = 4)
levels(datos$esite)=c("C3","C4","CZ","F3","F4","F7","F8","FP1","FP2","FZ","O1",
                      "O2","P3","P4","PZ","T3","T4","T5","T6", "VEOG")

#####
# Creación de la base de datos, determinando el paciente y
# el tipo de estímulo que se quiere estudiar
#####

paciente11<-(datos$subject==2)&(datos$gender=="fema")&(datos$wordtype=="1")
              &(datos$esite!="VEOG")
BASEPACIENTE1 = as.matrix(datos[paciente11,4:353])
basepaciente1 = t(BASEPACIENTE1)
dim(basepaciente1)

paciente12<-(datos$subject== 2)&(datos$gender=="fema")&(datos$wordtype=="2")
              &(datos$esite!="VEOG")
BASEPACIENTE2 = as.matrix(datos[paciente12,4:353])
basepaciente2 = t(BASEPACIENTE2)

#####
# Creación de la Matriz de distancias
#####

coord <-matrix(scan("CoorFinalesF.txt",0, dec=",",), 19, 2, byrow=TRUE)
plot(coord[,1], coord[,2])
dista <- dist(coord, method="euclidean", diag=TRUE, upper=TRUE)
d <- as.matrix(dista)

#####
# Suavizado de los datos
#####
coord.cero <- matrix(c(0,0),nrow=1,ncol=2)
n<-dim(basepaciente1)[1]
time<- c( seq(4,1400,4))
nbasis = 50
norder = 4

```

```
data=basepaciente1
data2=basepaciente2
coords=coord
nbasis=nbasis
argvals=time
max.dist = max(d)
smooth.type = "bsplines"
lambda = 0

s <- dim(data)[2]
  fdobjects <- .create.fd.object(data, smooth.type, argvals,
    nbasis, lambda)
  fd2objects <- .create.fd.object(data2, smooth.type, argvals,
    nbasis, lambda)

#####
# Estimación de las traza variogramas
#####

#####
# Cálculo de la norma L2
#####

L2_1 = l2.norm.cross(s, fdobjects$datafd,fdobjects$datafd, fdobjects$M)
L2_12 = l2.norm.cross(s, fdobjects$datafd,fd2objects$datafd, fdobjects$M)
L2_2 = l2.norm.cross(s, fd2objects$datafd,fd2objects$datafd, fd2objects$M)

#####
# Estimación de la traza variograma could
#####
okfd.res.cross11<-okfdLeft.crossfm(new.coords=coord.cero, coords=coord,
  data=basepaciente1,data2=basepaciente1, cov.model =NULL,
  nbasis=nbasis, argvals=time, fix.nugget=TRUE, kappa=0.7)

okfd.res.cross12<-okfdLeft.crossfm(new.coords=coord.cero, coords=coord,
  data=basepaciente1,data2=basepaciente2, cov.model = NULL,
  nbasis=nbasis, argvals=time,fix.nugget=TRUE, kappa=0.7)

okfd.res.cross22<-okfdLeft.crossfm(new.coords=coord.cero, coords=coord,
  data=basepaciente2,data2=basepaciente2, cov.model = NULL,
  nbasis=nbasis, argvals=time, fix.nugget=TRUE, kappa=0.7)

#####
# Estimación de la traza variograma bin
#####
```

```

trace1 = trace.vari1(coords, L2_1, bin=T ,uvec =10 , breaks = "default")
trace1$np
plot(trace1$u, trace1$v, col = 2)

trace2 = trace.vari1(coords, L2_2, bin=T, uvec = 9, breaks = "default")
trace2$np
plot(trace2$u, trace2$v, col = 2)

trace12 = trace.vari1(coords, L2_12, bin=T, uvec = 9, breaks = "default")
trace12$np
plot(trace12$u, trace12$v, col = 2)

#####
# Estimación de la traza variograma (bin y cloud)
#####
par(mfrow = c(1, 3))

variog1 = okfd.res.cross11$emp.trace.vari
plot(variog1, main = "Positivo", xlab = "distancia", ylab = "Semivarianza")
points(trace1$u, trace1$v, col = 2, lwd = 5)
variog1 = cbind(trace1$np, trace1$u, trace1$v,1)

variog2 = okfd.res.cross22$emp.trace.vari
plot(variog2, , main = "Negativo", xlab = "distancia", ylab = "Semivarianza")
points(trace2$u, trace2$v, col = 2, lwd = 5)
variog2 = cbind(trace2$np, trace2$u, trace2$v,2)

variog12 = okfd.res.cross12$emp.trace.vari
plot(variog12, , main = "Positivo - Negativo", xlab = "distancia",
      ylab = "Semivarianza")
points(trace12$u, trace12$v, col = 2, lwd = 5)
variog12 = cbind(trace12$np, trace12$u, trace12$v,3)

variogT = rbind(variog1,variog2,variog12)
v = variogT

#####
# Ajuste de las traza variogramas a través del M.L.C.
#####

h = seq(0,200,1)
modelo = vgm()
model=vgm(7000,"Exp", 100,fix.nugget = 0)

res = ejemploNN(v, model, fit.ranges = F, fit.lmc = TRUE, correct.diag = 1.01)
g1 = res
g1

```

```

par(mfrow = c(1, 3))
plot(okfd.res.cross11$emp.trace.vari, main = "Positivo", xlab = "distancia",
     ylab = "Semivarianza")
modelovariog1 = g1[1,1]*(exp(-d/g1[1,2]))
pmodelovariog1 = g1[1,1] - g1[1,1]*(exp(-h/g1[1,2]))
lines(h,pmodelovariog1)
points(trace1$u, trace1$v, col = 2, lwd = 5)
plot(okfd.res.cross22$emp.trace.vari, main = "Negativo", xlab = "distancia",
     ylab = "Semivarianza")
modelovariog2 = g1[2,1]*(exp(-d/g1[2,2]))
pmodelovariog2 = g1[2,1]- g1[2,1]*(exp(-h/g1[2,2]))
lines(h,pmodelovariog2)
points(trace2$u, trace2$v, col = 2, lwd = 5)
plot(okfd.res.cross12$emp.trace.vari, main = "Positivo - Negativo",
     xlab = "distancia", ylab = "Semivarianza")
modelovariog12 = g1[3,1]*(exp(-d/g1[3,2]))
pmodelovariog12 = g1[3,1]-g1[3,1]*(exp(-h/g1[3,2]))
lines(h,pmodelovariog12)
points(trace12$u, trace12$v, col = 2, lwd = 5)

#####
# Estimación de la matriz de varianzas y covarianzas
#####
f1 = cbind(modelovariog1,modelovariog12)
f2 = cbind(modelovariog12,modelovariog2)
W = rbind(f1,f2)
eigen(W)$values
solve(W)
Sigma2 = W
I38 = diag(38)

#####
# Código para determinar los parámetros funcionales
# de un ANOVAF con errores correlacionados
#####
#####
# Expresión de los datos y de los parámetros como objetos funcionales para
# comparar estímulos (positivo versus negativo).
#####

BASETOTALP15 = (datos$subject== 2)&(datos$gender=="fema")&((datos$wordtype=="1")|
                 (datos$wordtype=="2"))&(datos$esite!="VEOG")

BASETOTALP15 =as.matrix(datos[BASETOTALP15,4:353])
BASETOTALTP15 = t(BASETOTALP15)
t = dim(BASETOTALTP15)[1]
n = dim(BASETOTALTP15)[2]

```

```

time<- c( seq(4,1400,4))
rangeval = c(4,1400)
nbasis = 50
norder = 4
basis50 = create.bspline.basis(rangeval, nbasis, norder)
dist = max(d)*0.75
model = "exponential"
smooth.type = "bsplines"
lambda = 0

s <- dim(BASETOTALTP15)[2]
  fdobjects <- .create.fd.object(BASETOTALTP15, smooth.type, time,
    nbasis, lambda)
fd15h = data2fd(BASETOTALTP15, time, basis50)
plot(fd15h)

#####
# Creación de matrices de productos integrados.
#####
J = fdobjects$M
J = inprod(fdobjects$datafd,fdobjects$datafd)
J1 = inprod(fd15h$basis,fd15h$basis)

#####
# Creación de la matriz diseño.
#####

estimulos <- c("Efecto global", "Positivo ", "Negativo")

#####
#Indices para los estimulos
#####
p = 3
posiindex <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19)
negindex <- c(20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38)

#####
#Matriz diseño
#####

zmat <- matrix(0,n,p)
zmat[,1] <- 1
zmat[posiindex,2] <- 1
zmat[negindex,3] <- 1

#####

```

```
#Generación de la fila, para garantizar la ortogonalidad
#y estimabilidad de los parámetros.
#Fila (0,1,1,1)
#####

z39 <- matrix(1,1,3)
z39[1] <- 0
zmat39 <- rbind(zmat, z39)

#####
#Estimación de la fila generada a través de las
#funciones base.
#Fila (0,1,1).
#####

coef <- fd15h$coefs
coef39 <- cbind(coef,matrix(0,nbasis,1))
fd15h$coefs <- coef39
C = fd15h$coefs
#####
# Estimación de los parametros funcionales
#####
W = Sigma2
v39 = rep(0,n)
W1 = cbind(W,v39)
V239 = c(rep(0, n), 1)
W2 = rbind(W1,V239)
B39 = solve(t(zmat39)%*%solve(W2)%*%zmat39)%*%t(zmat39)%*%solve(W2)%*%t(C)
evalbasis<-eval.basis(time,basis50)
beta39 = B39%*%t(evalbasis)

par(mfrow = c(1, 2))
plot(time,beta39[1,], type = "l", col = 1, ylab = "Voltios",
      xlab = "Milisegundos", ylim = c(-8,10))
lines(time,beta39[2,], type = "l", col = 2)
lines(time,beta39[3,], type = "l", col = 3)
legend(100,-3,c("Media global","Positivo", "Negativo"), col = c(1,2,3),
      cex = 0.7, lty = 2, box.col = "white")

dim(beta39)

#####
# Estimación de las medias funcionales
#####
estimulo1 = beta39[1,] + beta39[2,]
estimulo2 = beta39[1,] + beta39[3,]
plot(time,estimulo1, type = "l", col = 2, ylim = c(-8,12),
```



```

      ylab = "Voltios", xlab = "Milisegundos")
lines(time,estimulo2, type = "l", col = 3)
legend(0,12,c("Positivo", "Negativo"), col = c(2,3), cex = 0.7,
      lty = 2, box.col = "white")

#####
# Estimación de SOLO la media
#####
MatrizMod = matrix(rep(1), n,1)
Cred = C[-(n+1)]
Bred = solve(t(MatrizMod)%*%solve(W)%*%MatrizMod)%*%t(MatrizMod)%*%solve(W)%*%t(Cred)
evalbasis<-eval.basis(time,basis50)
betared = Bred%*%t(evalbasis)
plot(time,betared[1,], type = "l", col = 2)

#####
# Código para calcular la prueba F propuesta por
# Shen and Faraway (2004) de un ANOVA con errores
# correlacionados
#####

#####
# Cálculo de la matriz de residuales y obtención de
# la matriz de varianza y covarianza
#####
Estimados = matrix(rep(0),n,t) # Valores estimados
for(i in 1:(n/2))
{
  Estimados[i,] = beta39[1,] + beta39[2,]
}

for(i in ((n/2)+1):38)
{
  Estimados[i,] = beta39[1,] + beta39[3,]
}
Estimados = t(Estimados)
ymat      <- eval.fd(time, fd15h)          # Valores observados
matplot(time,ymat[,1:n], type="l")
alfa <- betared                          # Constante (Media) estimada
plot(time,alfa, xlab="Tiempo", ylab="Voltaje", type = "l")
Residuales <- ymat[,1:n] - Estimados[,1:n] # Los residuales son iguales a
                                           # observados - estimados
SUMACUADRAMCGCOM = Residuales%*%solve(W)%*%t(Residuales)
matplot(Residuales, type = "l")

#####
#####

```

```

ymat = ymat[,1:n]
Y = t(ymat)
Estimados = t(Estimados)

#####
#Matriz diseño Modelo completo
#####
p = 3
zmat <- matrix(0,n,p)
zmat[,1] <- 1
zmat[posiindex,2] <- 1
zmat[negindex,3] <- 1

#####
# Matriz de suma de cuadrados de los residuales
# del modelo completo
#####

MatrizResCom = (t(Y-zmat%*%beta39))%*%solve(W)%*%(Y-zmat%*%beta39)
dim(MatrizResCom)
sum(diag(MatrizResCom))
#####
#Matriz diseño Modelo reducido
#####
q = 1
zmatRed <- matrix(0,n,q)
zmatRed[,1] <- 1

#####
# Matriz de suma de cudarados de los residuales
# del modelo reducido
#####

MatrizResRed = (t(Y-zmatRed%*%betared))%*%solve(W)%*%(Y-zmatRed%*%betared)
dim(MatrizResRed)
sum(diag(MatrizResRed))
#####
# Prueba F de Fareway: MCG
#####

NumF = (sum(diag(MatrizResRed-MatrizResCom))/(p-q))
DenF = ((sum(diag(MatrizResCom))/(n-p)))
FCalculada = (NumF/DenF)
factorPon = (((sum(diag(MatrizResCom)))^2)/ (sum(diag(MatrizResCom^2))))
f1Pon = factorPon*(p-q)
f2Pon = factorPon*(n-p)
FteoPon<-qf(0.95,f1Pon,f2Pon)
FCalculada

```

```
#####
# Código para realizar la prueba propuesta por
# Cuestas-Albertos and Febrero-Bande (2007) de
# un ANOVA con errores correlacionados
#####
#####
# Generación del movimiento Browniano (v)
#####
Z = rnorm(350,0,1)
h = seq(1,350,4)
t = matrix(rep(0),350,1)
x = matrix(rep(0),19,1)
y = matrix(rep(0),19,1)
for(i in 1:350)
{
  t[i] = 4*h[i]
}
v = matrix(rep(0),350,1)
for(i in 2:350)
{
  v[i] = v[i-1]+ 4*Z[i]
}

#####
# Cálculo de las proyecciones de las curvas
# sobre el espacio generado por v
#####
for(i in 1:19)
{
  vectorGranx = cbind(basepaciente1[,i],v)
  vectorGranx = as.matrix(vectorGranx)
  vectorGrany = cbind(basepaciente2[,i],v)
  vectorGrany = as.matrix(vectorGrany)

  n<-dim(vectorGranx)[2]
  time<- c( seq(4,1400,4))
  nbasis = 50
  norder = 4
  smooth.type = "bsplines"
  lambda = 0
  vectorFx= .create.fd.object(vectorGranx, smooth.type, time, nbasis, lambda)
  vectorFy= .create.fd.object(vectorGrany, smooth.type, time, nbasis, lambda)
  x[i] = 12.norm.cross(n,vectorFx$dataafd,vectorFx$dataafd, vectorFx$M)[1,2]
  y[i] = 12.norm.cross(n,vectorFy$dataafd,vectorFy$dataafd, vectorFy$M)[1,2]
}

#####
```

```
# Estimación de la matriz de varianza y covarianza
# mediante la estimación y el ajuste de la semivarianza
# y la semivarianza cruzada
#####
dista <- dist(coord, method="euclidean", diag=TRUE, upper=TRUE)
d <- as.matrix(dista)
n = 19
coord
n.tiempo<- 2
mean<-cbind(coord,x,y)
#####
# Ajuste de las sevivarianzas mediante M.L.C
#####
meanc<-as.data.frame(mean)
n2<-paste("var",1:n.tiempo, sep="")
names(meanc)<-c("x","y",n2)
coordinates(meanc)= ~x+y

g<-NULL
for (i in 1:n.tiempo)
{
  g <- gstat(g,formula= as.formula(paste(n2[i],"~1")), data=meanc)
}
variog <- variogram(g)
plot(variog)
g <- gstat(g, model=vgm(7000,"Gau", 100, fix.nugget = 0), fill.all=TRUE)
g.fit <- fit.lmc(variog,g,fit.lmc=TRUE)
plot(variog,g.fit)

modelovariog1 = g.fit$model$var1[1,2]*(exp(-d/g.fit$model$var1[1,3]))
modelovariog2 = g.fit$model$var2[1,2]*(exp(-d/g.fit$model$var2[1,3]))
modelovariog12 =g.fit$model$var1.var2[1,2]*(exp(-d/g.fit$model$var1.var2[1,3]))

#####
# Estimación de la matriz de varianza y covarianza
#####

f1 = cbind(modelovariog1,modelovariog12)
f2 = cbind(modelovariog12,modelovariog2)
W = rbind(f1,f2)
fila = rep(0,38)
W1 = cbind(W,fila)
V239 = c(rep(0, 38), 1)
W2 = rbind(W1,V239)

#####
# Establecimiento de la matriz del modelo completo
```

```
#####

Y = rbind(x,y,0)
MatrizModelo = matrix(rep(1),38,3)
MatrizModelo[,2] = rbind(matrix(rep(1),19,1),matrix(rep(0),19,1))
MatrizModelo[,3] = rbind(matrix(rep(0),19,1),matrix(rep(1),19,1))
fila = c(0,1,1)
MatrizModelo2 = rbind(MatrizModelo,fila)
#####
# Establecimiento de la matriz del modelo reducido
#####

MatrizModeloRed = matrix(rep(1),38,1)

#####
# Estimación de los parámetros del ANOVA clásico con
# errores correlacionados
#####

Beta = solve(t(MatrizModelo2)%*%solve(W2)%*%MatrizModelo2)%*%
      t(MatrizModelo2)%*%solve(W2)%*%Y
Y =Y[1:38,]
BetaRed = solve(t(MatrizModeloRed)%*%solve(W)%*%MatrizModeloRed)%*%
          t(MatrizModeloRed)%*%solve(W)%*%Y

#####
# Estimación de la suma de cuadrados de los residuales y
# de la prueba F clásica con errores correlacionados
#####

SCR = t(Y-MatrizModelo%*%Beta)%*%solve(W)%*%(Y-MatrizModelo%*%Beta)
SCRred = t(Y-MatrizModeloRed%*%BetaRed)%*%solve(W)%*%(Y-MatrizModeloRed%*%BetaRed)
p = 3
q = 1
n = 38
FNum = (SCRred-SCR)/(p-q)
FDen = SCR/(n-p)
F = FNum/FDen
F
```

Bibliografía

- Ashburner, J. and Friston, K. (1997). *Spatial transformation of images*. Human brain function.
- Bauer, L. and Hesselbrock, V. (1999). P300 decrements in conduct disorder in teenagers: Implications for substance abuse risk and brain development. *Biol Psychiatry*, 46.
- Borovskikh, Y. (1996). *U-statistics in Banach spaces*. Utrecht VSP.
- Chatrian, G. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked eeg activity. *Am J EEG Technol*, 25:83–92.
- Cressie, N. (1993). *Statistic for spatial data*. John Wiley Sons.
- Cuesta-Albertos, J. and Febrero-Bande, M. (2007). A simple multiway anova for functional data. *Business and Economics*, 19:537–557.
- Cuevas, A. Febrero, M. and Fraiman, R. (2004). An anova test for functional data. *Computational Statistics and Data Analysis*, 47:111–122.
- Delicado, P. (2007). Functional k-sample problem when data are density functions. *Computational Statistics*, 22 (3):391–410.
- Draper, N. and Smith, H. (1981). *Applied Regression Analysis*. Wiley Series in Probability and Statistics.
- Duffy, F., B. P. and Burchfiel, J. (1981). Significance probability mapping: An aid in the topographic analysis of brain electrical activity. *Electroencephalography and Clinical Neurophysiology*, 51:455–462.
- Ferraty, F. and Vieu, P. (2006). *Non parametric functional data analysis. Theory and practice*. Springer.
- Friston, K., A. J. K.-E. N. T. and Penny, W. (2006). *Statistical parametric mapping: The analysis of functional brain images*. Academic Press.
- Giraldo, R. (2009). *Geostatistics analysis for functional data*. PhD thesis, Universitat Politècnica de Catalunya.

- González, P. (2009). Potenciales evocados cognitivos (p300) en dolor oncológico crónico. *Revista Mexicana de Medicina Física y Rehabilitación*, 21:78–83.
- Goulard, M. and Voltz, M. (1991). Geostatistical interpolation of curves: A case study in soil science. *Geostatistics Troia 92*, 2:805–816.
- Gratton, G. (1998). Dealing with artifacts: The eeg contamination of event-related brain potential. *Research Methods, Instruments and Computers*, 30:44–53.
- Harmony, T., F. T. F.-B. A. S.-P. J. B. J. D.-C. L. and Galán, L. (2001). Eeg changes during word and figure categorization.
- Hasper, H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol*, 10:371–375.
- Herscovitchet, P. (1983). Brain blood flow measured with intravenous h215. theory and error analysis. *J Nucl Med*, 24.
- Isaaks, E. and Srivastava, M. (1987). *Applied geostatistics*. Oxford University Press.
- Lauter, J. (1985). *Tonotopic organization in human auditory contex revealed by positron emission tomography*. Hearing Research.
- Lueck, C., Z. S. F.-K. D. M.-C. P. C. V. L.-A. K. C. and Frackowiak, R. (1989). *Colour centre in the cerebral cortex of man*. Nature.
- Oostenveld, R. and Praamstra, P. (2001). The five percent electrode system for high-resolution eeg and erp measurements. *Clin Neurophysiol*, 112:713–719.
- Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society*, 3:539–572.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer.
- Shen, Q. and Faraway, J. (2004). An f test for linear models with functional responses. *Statistica Sinica*, 14:1239–1257.
- Sánchez, P. (1992). Métodos para el procesamiento y análisis estadístico multivariante de señales multicanal. aplicación al estudio eeg. Master's thesis, Universidad de la Laguna.
- Stein, P. (1999). Interpolation of spatial data. some theory of kriging. *Springer*.
- Stuart, C. and Robert, E. (2008). *Neuropsychiatry and behavioral neurosciences*. The American Psychiatric Publishing.
- Worsley, K., P. J. F.-K. and Evans, A. (1997). Characterizing the response of pet and fmri data using multivariate linear models. *NeuroImage*, 6:305–319.